AMERICAN SOCIETY FOR MICROBIOLOGY | Antimicrobial Agents and Chemotherapy®

MECHANISMS OF RESISTANCE

# A Standard Numbering Scheme for Class C β-Lactamases

Andrew R. Mack,[a,f] Melissa D. Barnes,[b,f] Magdalena A. Taracila,[b,f] Andrea M. Hujer,[b,f] Kristine M. Hujer,[b,f] Gabriel Cabot,[w,x] Michael Feldgarden,[g] Daniel H. Haft,[g] William Klimke,[g] Focco van den Akker,[d] Alejandro J. Vila,[l,n,o,p] Andrea Smania,[q,r] Shozeb Haider,[dd] Krisztina M. Papp-Wallace,[b,d,e,f] Patricia A. Bradford,[h] Gian Maria Rossolini,[s,t] Jean-Denis Docquier,[ee] Jean-Marie Frère,[u] Moreno Galleni,[u] Nancy D. Hanson,[ff] Antonio Oliver,[w,x] Patrick Plésiat,[y,z,aa] Laurent Poirel,[bb,cc] Patrice Nordmann,[bb,cc] Timothy G. Palzkill,[i,j] George A. Jacoby,[v] Karen Bush,[k] Robert A. Bonomo[a,b,c,d,e,f,l,m]

[a]Department of Molecular Biology and Microbiology, Case Western Reserve University, Cleveland, Ohio, USA
[b]Department of Medicine, Case Western Reserve University, Cleveland, Ohio, USA
[c]Department of Pharmacology, Case Western Reserve University, Cleveland, Ohio, USA
[d]Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio, USA
[e]Department of Proteomics and Bioinformatics, Case Western Reserve University, Cleveland, Ohio, USA
[f]Research Service, Louis Stokes Cleveland Department of Veterans Affairs, Cleveland, Ohio, USA
[g]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA
[h]Antimicrobial Development Specialists, LLC, Nyack, New York, USA
[i]Department of Pharmacology and Chemical Biology, Baylor College of Medicine, Houston, Texas, USA
[j]Verna Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas, USA
[k]Department of Biology, Indiana University Bloomington, Bloomington, Indiana, USA
[l]CWRU-Cleveland VAMC Center for Antimicrobial Resistance and Epidemiology (Case VA CARES), Cleveland, Ohio, USA
[m]Geriatric Research Education and Clinical Centers (GRECC), Louis Stokes Cleveland Department of Veterans Affairs, Cleveland, Ohio, USA
[n]Instituto de Biología Molecular y Celular de Rosario (IBR, CONICET-UNR), Rosario, Argentina
[o]Plataforma Argentina de Biología Estructural y Metabolómica PLABEM, Buenos Aires, Argentina
[p]Área Biofísica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina
[q]Centro de Investigaciones en Química Biológica de Córodoba (CIQUIBIC), CONICET, Universidad Nacional de Córdoba, Córdoba, Argentina
[r]Departamento de Química Biológica Ranwel Caputto, Facultad de Ciencias Químicas, Universidad Nacional de Córdoba, Córdoba, Argentina
[s]Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy
[t]Microbiology and Virology Unit, Florence Careggi University Hospital, Florence, Italy
[u]Centre for Protein Engineering (CIP), University of Liège, Liège, Belgium
[v]Lahey Hospital and Medical Center, Burlington, Massachusetts, USA
[w]Red Española de Investigación en Patología Infecciosa (REIPI), Instituto de Salud Carlos III, Madrid, Spain
[x]Servicio de Microbiología, Hospital Universitario Son Espases, Instituto de Investigación Sanitaria Illes Balears (IdISBa), Palma de Mallorca, Spain
[y]Laboratoire de Bactériologie, Centre Hospitalier Régional Universitaire, Besançon, France
[z]UMR6249 CNRS Chrono-Environnement, Université de Bourgogne Franche-Comté, Besançon, France
[aa]Centre National de Référence de la Résistance aux Antibiotiques, Besançon, France
[bb]Emerging Antibiotic Resistance Unit, Medical and Molecular Microbiology, Faculty of Science and Medicine, University of Fribourg, Fribourg, Switzerland
[cc]Swiss National Reference Center for Emerging Antibiotic Resistance (NARA), University of Fribourg, Fribourg, Switzerland
[dd]UCL School of Pharmacy, University College London, London, United Kingdom
[ee]Department of Medical Biotechnology, University of Siena, Siena, Italy
[ff]Department of Microbiology and Immunology, Creighton School of Medicine, Omaha, Nebraska, USA

**ABSTRACT** Unlike for classes A and B, a standardized amino acid numbering scheme has not been proposed for the class C (AmpC) β-lactamases, which complicates communication in the field. Here, we propose a scheme developed through a collaborative approach that considers both sequence and structure, preserves traditional numbering of catalytically important residues (Ser$^{64}$, Lys$^{67}$, Tyr$^{150}$, and Lys$^{315}$), is adaptable to new variants or enzymes yet to be discovered and includes a variation for genetic and epidemiological applications.

**Citation** Mack AR, Barnes MD, Taracila MA, Hujer AM, Hujer KM, Cabot G, Feldgarden M, Haft DH, Klimke W, van den Akker F, Vila AJ, Smania A, Haider S, Papp-Wallace KM, Bradford PA, Rossolini GM, Docquier J-D, Frère J-M, Galleni M, Hanson ND, Oliver A, Plésiat P, Poirel L, Nordmann P, Palzkill TG, Jacoby GA, Bush K, Bonomo RA. 2020. A standard numbering scheme for class C β-lactamases. Antimicrob Agents Chemother 64:e01841-19. https://doi.org/10.1128/AAC.01841-19.



Address correspondence to Robert A. Bonomo, robert.bonomo@va.gov.

For a commentary on this article, see https://doi.org/10.1128/AAC.02247-19.

**Received** 9 September 2019
**Returned for modification** 29 September 2019
**Accepted** 8 November 2019
**Accepted manuscript posted online** 11 November 2019
**Published** 21 February 2020

**KEYWORDS** AmpC, amino acid numbering, beta-lactamases, class C beta-lactamase, conserved residue, nomenclature, structure-activity relationships

An urgent need exists to address current inconsistencies in the numbering of amino acid residues among class C $\beta$-lactamases, both within families and across the class. Established conventions in the field define three common features shared among the serine-type $\beta$-lactamases. In the class C $\beta$-lactamases, also known as AmpC $\beta$-lactamases, these features occur at recognizable conserved motifs: $S^{64}$XXK (where $S^{64}$ is the active-site serine), $Y^{150}$XN, and $K^{315}$(S/T)G (1–4). These designations align with the amino acid sequence of the mature form of both the P99 AmpC (originally characterized in an *Enterobacter cloacae* strain [NCBI RefSeq accession no. WP_049134845.1] but now found to be an *Enterobacter hormaechei* strain [GenBank accession no. CAA30257.1]) and *Escherichia coli* AmpC (NCBI RefSeq accession no. WP_001336292.1; originally characterized in strain K-12 under GenBank accession no. AAC77110.1). While *E. coli* AmpC has historical significance as both the first $\beta$-lactamase reported (5) and the first class C $\beta$-lactamase sequenced, P99 maintains the same numbering of conserved motifs and the mature form begins with residue one (6) rather than residue four as in *E. coli* AmpC. In contrast, many other class C $\beta$-lactamases possess insertions and deletions that shift the numbering of the conserved residues, significantly complicating both nomenclature and comparisons between enzymes.

For this report, we analyzed 155 unique AmpC structures deposited in the Protein Data Bank (including 142 supported by 66 publications) and found that 129 $\beta$-lactamase structures identify the catalytic serine as $Ser^{64}$ (123 naturally and 6 with alignment), 10 number from the beginning of the precursor form with the signal peptide included, and the remaining 16 number from the beginning of the mature form but do not identify the catalytic serine as $Ser^{64}$ (of which 8 are not associated with a publication). Additionally, based on a literature search of PubMed, we found that consistency is lacking for numbering within the various families of class C $\beta$-lactamases. As an example, since the term PDC (*Pseudomonas*-derived cephalosporinase) was coined in 2009 for the chromosomal AmpC of *Pseudomonas aeruginosa*, three different approaches have been used to number amino acid residues in this $\beta$-lactamase (7). These approaches include (i) direct numbering of residues beginning with the N terminus of the precursor protein (7), (ii) direct numbering of residues beginning with the N terminus of the mature protein (8), and (iii) alignment-based numbering designed to maintain the conventional assignment of conserved residues and to simplify numbering for comparisons across families (9). Unfortunately, it can be unclear to readers which of the various schemes is being used in a given publication. As a result, authors may sometimes find choosing a numbering scheme and numerically designating a given residue problematic. Comparing findings from multiple publications may be made unnecessarily difficult; resolving ambiguity in assignment may be extremely challenging. For example, a reference to Gly at position 183 in PDC may refer to a site that is described as having a clinically relevant mutation if numbering begins with $Met^{1}$ of the precursor form but would refer to a different glycine, 26 residues away, if alignment-based numbering was used (10, 11).

To address this growing concern, we propose a numbering scheme to use consistently when referring to crystallographically equivalent positions in the mature form of any class C $\beta$-lactamase. We suggest the acronym "SANC" to name the scheme, for structural alignment-based numbering of class C $\beta$-lactamases, or else the simpler term "structural position." In developing this numbering scheme, we adapted the approaches used by Ambler et al. for the class A $\beta$-lactamases (12) and Galleni et al. for the class B $\beta$-lactamases (13). We conducted an amino acid alignment of 32 AmpC $\beta$-lactamases, both chromosomal and plasmid encoded (see Table S1 in the supplemental material) and identified characteristic differences from P99 for each enzyme (Table 1). Sequences were obtained from the National Center for Biotechnology Information Protein Database (14), and signal peptide cleavage

**TABLE 1** Insertions and deletions present in the AmpC enzymes examined in comparison to *E. cloacae* complex P99

| Class C β-lactamase | NCBI accession no. | Insertions and deletions relative to *E. cloacae* complex P99[a] |
|---|---|---|
| ACC-1 | WP_032491956.1 | −116, +204a, +247a, −289, −290, +362, +363 |
| ACT-1 | WP_063857727.1 | −361 |
| ADC-7 | WP_063857816.1 | +0, +204a, −245, −304, −305, −306, +362 |
| ADC-8 | WP_004923134.1 | +0d, +0c, +0b, +0a, +0, −245, +362, +363, +364, +365, +366, +367 |
| AQU-1 | WP_099156042.1 | −1, −2, +204a, −243, −245, −301, −302, +362 |
| *Burkholderia multivorans* AmpC1 | WP_012218336.1 | +204a, −245, +362, +363, +364 |
| BUT-1 | WP_104531863.1 | +0a, +0 |
| CepH | WP_063843234.1 | −1, −2, +204a, −243, −245, +362 |
| CepS | WP_063843235.1 | −1, −2, +204a, −243, −245, +362 |
| CFE-1 | WP_032490699.1 | None |
| CMA-1 | WP_032974004.1 | −1, −2, −3, −4, −5, −6, −116, +204a, −245, +362 |
| CMH-1 | WP_063859580.1 | None |
| CMY-2 | WP_000976514.1 | None |
| CSA-1 | WP_007888761.1 | −1, −2, −3, −4, −5, −6, −116, +204a, −245, +362 |
| DHA-1 | WP_004236386.1 | −1, −2, −3, −4, −301 |
| *Escherichia coli* AmpC | WP_001336292.1 | −1, −2, −3 |
| EC-5 | WP_001443153.1 | −1, −2, −3 |
| *Edwardsiella* AmpC | WP_041692555.1 | +0a, +0 |
| FOX-4 | WP_032489727.1 | −1, −2, +204a, −243, −245, +362 |
| LHK-1 | WP_081666691.1 | −1, −2, −3, −4, +204a, −245 |
| LRA-10 | WP_099982803.1 | −1, −126, +204a, −245, −361 |
| LRA-18 | WP_099982801.1 | −1, −245, −311, +362, +363, +364, +365 |
| *Mycobacterium smegmatis* AmpC | WP_011729443.1 | −1, −2, −3, −4, −5, −6, +204a, −245, −305, −306, +362 |
| MIR-1 | WP_032489464.1 | None |
| MOX-1 | WP_032489888.1 | +0, +204a, −243, −245, −301, −302, −303, +362 |
| OCH-1 | WP_040129485.1 | +0, +204a, −245, +362, +363, +364 |
| PAC-1 | WP_034051940.1 | −1, −2, −3, −4, −5, −116, +204a, −245, +362 |
| PDC-1 | WP_003101289.1 | +125a, +204a, −245, +362, +363, +364, +365, +366, +367, +368, +369, +370 |
| SRT-1 | WP_063864749.1 | −1, −2, −3, −4, −5, −6, −116, +204a, −245, +362, +363 |
| SST-1 | WP_063864750.1 | −1, −2, −3, −4, −5, −6, −116, +204a, −245, +362, +363 |
| TRU-1 | WP_042027926.1 | −1, −2, +204a, −243, −245, +362 |

[a]A minus indicates a deletion, and a plus indicates an insertion. Appended letters indicate that an insertion follows a given residue number.

sites were determined using UniProt (or SignalP 5.0 for entries not present in UniProt) (15, 16). Mature protein sequences were aligned using the MUSCLE algorithm (17) with default settings.

Consensus secondary structure (defined as a majority of structures in agreement for a given amino acid position) was determined based on comparisons of a representative structure of each of the 10 AmpC β-lactamases for which one or more structures are available in the Protein Data Bank, specifically, ACT-1 (PDB ID 2ZC7), ADC-7 (PDB ID 4U0T), CMY-2 (PDB ID 1ZC2), *E. coli* AmpC (PDB ID 2BLS), FOX-4 (PDB ID 5CGS), MOX-1 (PDB ID 3W8K), *Mycobacterium smegmatis* AmpC (PDB ID 5E2H), PDC-1 (PDB ID 4GZB), and TRU-1 (PDB ID 6FM6). The consensus agrees with the secondary structure (or lack thereof) of P99 for just over 90% of residues. This consensus was used to annotate secondary structure, including stripes to indicate residues with an even split between two secondary structure types, and helix numbers on the alignment. Finally, a simple literature survey was conducted to determine residues belonging in either the consensus portion or the fullest likely extent of the Ω-loop or R2-loop, both of which are also annotated on the alignment. By including this structural information, we hope to both better correlate the numbering system with well-known structural features and to provide additional points of reference for those just beginning to work with AmpC structures.

The exact positions of one insertion and one deletion within the alignment were manually adjusted (residue 203a by MUSCLE became 204a by structure to preserve a β-turn, and the deletion of residue 247 by MUSCLE became a deletion of residue 245 by structure to preserve an α-helix) to ensure that they occurred in structurally reasonable areas of both the consensus structure and 10 source structures.

Amino acid numbering was based on *E. cloacae* complex P99 while preserving the conventional numbering of the following residues: $Ser^{64}$, $Lys^{67}$, $Tyr^{150}$, and

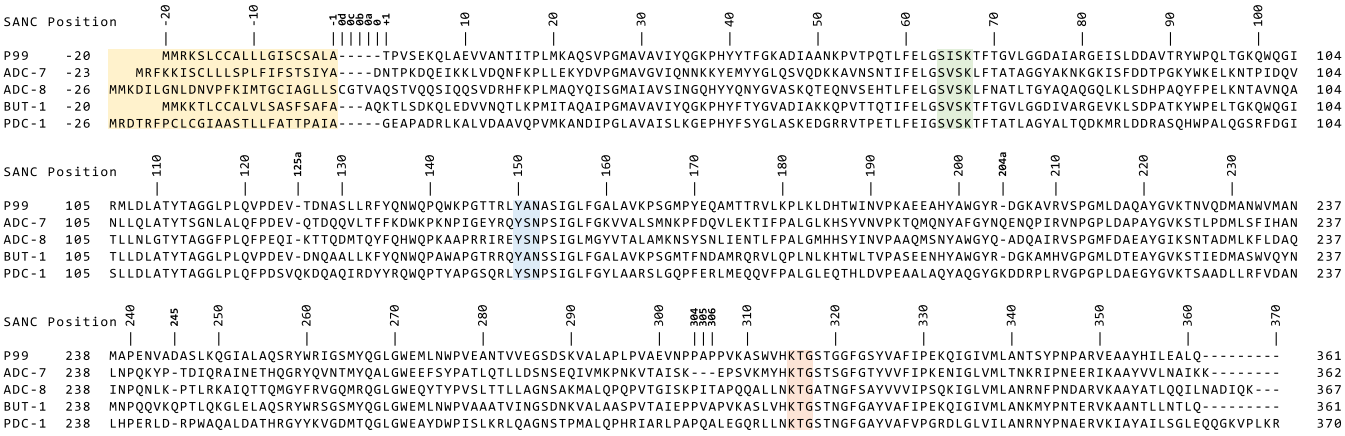


**FIG 1** Use of alignment to assign SANC-based amino acid residue numbers. Positions corresponding to insertions and deletions are indicated in bold. ADC-7 adds residues 0, 204a, and 262 and deletes residues 245 and 304 to 306. ADC-8 adds residues 0 to 0d and 262 to 267 and deletes residue 245. BUT-1 adds residue 0. PDC-1 adds residues 125a, 204a, and 362 to 370 and deletes residue 245. For reference, signal sequences are highlighted in yellow, $S^{64}$XXK in green, $Y^{150}$XN in blue, and $K^{315}$(S/T)G in red.

Lys[315]. Insertions relative to P99 were addressed by appending a lowercase letter(s) to the number of the amino acid immediately preceding the insertion (e.g., 125a in PDC-1). Deletions relative to P99 were skipped, resulting in "ghost residues" (e.g., ACC-1 has residues G115 and L117 with a deleted residue at position 116). For mature enzymes with more C-terminal amino acid residues than P99, additional residues are assigned numbers in numerical order beginning with 362. For mature enzymes with more N-terminal amino acid residues than P99, the first additional residue is numbered 0 and subsequent residues are numbered by appending a lowercase letter to zero while moving in an N-terminal direction (e.g., 0 and 0a for BUT-1 and *Edwardsiella* AmpC). Signal peptide residues are assigned negative numbers, beginning with −1 for the residue adjacent to the cleavage site and proceeding in the N-terminal direction until all residues are numbered. Multiple sequence alignments are not considered for the signal peptide regions. Figure 1 illustrates these principles with several examples.

Amino acid positions should be provided under both a family-specific, precursor-based scheme (precursor numbering) and the alignment-based scheme (SANC) at first mention of a given residue in a publication. Authors are free to choose their favored convention for subsequent mentions, but as a general suggestion, we encourage the use of SANC for biochemical and structural publications and precursor numbering for genetic and epidemiological publications.

Providing numbering under both schemes is essential to our proposal. Structural numbering maintains continuity with the conventional assignment of the catalytic serine as Ser[64] and the majority of existing literature on class C $\beta$-lactamase structure and function, while precursor numbering enables direct gene translation and simplifies interpretation of sequencing results, particularly within a single family. Utilizing this hybrid approach, an initial description of a typical PDC variant might read "PDC-221 differs from PDC-1 (GenBank accession no. AAG07497.1) by a single amino acid substitution, E247K, occurring at SANC position 219."

In Table S1 in the supplemental material, we show a multiple sequence alignment of 32 class C $\beta$-lactamases with column headers indicating the appropriate number to be used at each position. The spreadsheet also features a text-based alignment of the structures used in determining the consensus secondary structure. In Text S2 in the supplemental material, we provide a protein profile hidden Markov model (HMM) which implements the SANC scheme, built from the multiple sequence alignment using HMMER (http://hmmer.org). Alignments of the HMM to class C $\beta$-lactamases are expected to produce correct SANC assignments when results of the search are examined. We suggest using the HMM, rather than examinations by eye, to make position

assignments under this scheme for novel AmpC enzymes that may be discovered in the future. Finally, basic instructions for using our HMM with the HMMER software are also included in Text S1 in the supplemental material.

For the specific case of PDC variants, a database utilizing the three numbering schemes (SANC and both family-specific precursor and mature form numbering) is freely available at https://arpbigidisba.com/pseudomonas-aeruginosa-derived-cephalosporinase-pdc-database/.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, XLSX file, 0.1 MB.

**SUPPLEMENTAL FILE 2**, PDF file, 0.5 MB.

**SUPPLEMENTAL FILE 3**, TXT file, 0.2 MB.

## ACKNOWLEDGMENTS

We thank Ram Podicheti and Cameron Divoky for sequencing and species verification of the P99-producing *Enterobacter hormaechei* strain originally provided by Mark Richmond.

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (NIH) to R.A.B. under award numbers R01AI100560, R01AI063517, and R01AI072219. This study was also supported in part by funds and/or facilities provided by the Cleveland Department of Veterans Affairs to K.M.P.-W. and R.A.B., Veterans Affairs Merit Review Program award number 1I01BX002872 to K.M.P.-W., Veterans Affairs Merit Review Program award number 1I01BX001974 to R.A.B. from the Biomedical Laboratory Research & Development Service of the VA Office of Research and Development, and by the Geriatric Research Education and Clinical Center VISN 10. The work of Michael Feldgarden, Daniel H. Haft, and William Klimke was supported by the Intramural Research Program of the National Library of Medicine, National Institutes of Health.

N.D.H. receives funding from Alimetrix, bioMérieux, Merck, Roche, and Shionogi and serves as a scientific advisor for Streck.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the U.S Department of Veterans Affairs, or the U.S. Government.

## REFERENCES

1. Bush K. 2013. The ABCD's of β-lactamase nomenclature. J Infect Chemother 19:549–559. https://doi.org/10.1007/s10156-013-0640-7.
2. Jacoby GA. 2009. AmpC β-lactamases. Clin Microbiol Rev 22:161–182. https://doi.org/10.1128/CMR.00036-08.
3. Oefner C, D'Arcy A, Daly JJ, Gubernator K, Charnas RL, Heinze I, Hubschwerlen C, Winkler FK. 1990. Refined crystal structure of β-lactamase from *Citrobacter freundii* indicates a mechanism for β-lactam hydrolysis. Nature 343:284–288. https://doi.org/10.1038/343284a0.
4. Lobkovsky E, Moews PC, Liu H, Zhao H, Frere JM, Knox JR. 1993. Evolution of an enzyme activity: crystallographic structure at 2-Å resolution of cephalosporinase from the *ampC* gene of *Enterobacter cloacae* P99 and comparison with a class A penicillinase. Proc Natl Acad Sci U S A 90:11257–11261. https://doi.org/10.1073/pnas.90.23.11257.
5. Abraham EP, Chain E. 1940. An enzyme from bacteria able to destroy penicillin. Nature 146:837. https://doi.org/10.1038/146837a0.
6. Galleni M, Lindberg F, Normark S, Cole S, Honore N, Joris B, Frere JM. 1988. Sequence and comparative analysis of three *Enterobacter cloacae* ampC β-lactamase genes and their products. Biochem J 250:753–760. https://doi.org/10.1042/bj2500753.
7. Rodríguez-Martínez J-M, Poirel L, Nordmann P. 2009. Extended-spectrum cephalosporinases in *Pseudomonas aeruginosa*. Antimicrob Agents Chemother 53:1766–1771. https://doi.org/10.1128/AAC.01410-08.
8. Berrazeg M, Jeannot K, Enguéné VYN, Broutin I, Loeffert S, Fournier D, Plésiat P. 2015. Mutations in β-lactamase AmpC increase resistance of *Pseudomonas aeruginosa* isolates to antipseudomonal cephalosporins. Antimicrob Agents Chemother 59:6248–6255. https://doi.org/10.1128/AAC.00825-15.
9. Drawz SM, Taracila M, Caselli E, Prati F, Bonomo RA. 2011. Exploring sequence requirements for $C_3/C_4$ carboxylate recognition in the *Pseudomonas aeruginosa* cephalosporinase: insights into plasticity of the AmpC β-lactamase. Protein Sci 20:941–958. https://doi.org/10.1002/pro.612.
10. MacVane SH, Pandey R, Steed LL, Kreiswirth BN, Chen L. 2017. Emergence of ceftolozane-tazobactam-resistant *Pseudomonas aeruginosa* during treatment is mediated by a single AmpC structural mutation. Antimicrob Agents Chemother 61:e01183-17. https://doi.org/10.1128/AAC.01183-17.
11. Cabot G, Bruchmann S, Mulet X, Zamorano L, Moyà B, Juan C, Haussler S, Oliver A. 2014. *Pseudomonas aeruginosa* ceftolozane-tazobactam resistance development requires multiple mutations leading to overexpression and structural modification of AmpC. Antimicrob Agents Chemother 58:3091–3099. https://doi.org/10.1128/AAC.02462-13.
12. Ambler RP, Coulson AF, Frère JM, Ghuysen JM, Joris B, Forsman M, Levesque RC, Tiraby G, Waley SG. 1991. A standard numbering scheme for the class A β-lactamases. Biochem J 276:269–270. https://doi.org/10.1042/bj2760269.
13. Galleni M, Lamotte-Brasseur J, Rossolini GM, Spencer J, Dideberg O, Frère J-M, The Metallo-β-Lactamase Working Group. 2001. Standard numbering scheme for class B β-lactamases. Antimicrob Agents Chemother 45:660–663. https://doi.org/10.1128/AAC.45.3.660-663.2001.
14. NCBI Resource Coordinators. 2018. Database resources of the National

Center for Biotechnology Information. Nucleic Acids Res 46:D8–D13. https://doi.org/10.1093/nar/gkx1095.
15. The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. https://doi.org/10.1093/nar/gky1049.
16. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423. https://doi.org/10.1038/s41587-019-0036-z.
17. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.