



HAL
open science

A Standard Numbering Scheme for Class C β -Lactamases

Andrew Mack, Melissa Barnes, Magdalena Taracila, Andrea Hujer, Kristine Hujer, Gabriel Cabot, Michael Feldgarden, Daniel Haft, William Klimke, Focco van den Akker, et al.

► **To cite this version:**

Andrew Mack, Melissa Barnes, Magdalena Taracila, Andrea Hujer, Kristine Hujer, et al.. A Standard Numbering Scheme for Class C β -Lactamases. *Antimicrobial Agents and Chemotherapy*, 2020, 64 (3), 10.1128/AAC.01841-19 . hal-03482894

HAL Id: hal-03482894

<https://univ-fcomte.hal.science/hal-03482894v1>

Submitted on 19 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **A standard numbering scheme for class C β -lactamases**

2 Andrew R Mack^{1,6}, Melissa D Barnes,^{2,6} Magdalena A Taracila^{2,6}, Andrea M Hujer^{2,6}, Kristine M Hujer
3^{2,6}, Gabriel Cabot^{24,25}, Michael Feldgarden⁷, Daniel H Haft⁷, William Klimke⁷, Focco van den Akker⁴,
4 Alejandro J Vila^{12,14-17}, Andrea Smania^{18,19}, Shozeb Haider³¹, Krisztina M Papp-Wallace^{2,4,5,6}, Patricia A
5 Bradford⁸, Gian Maria Rossolini^{20,21}, Jean-Denis Docquier³², Jean-Marie Frère²², Moreno Galleni²²,
6 Nancy D Hanson³³, Antonio Oliver^{24,25}, Patrick Plésiat^{26,27,28}, Laurent Poirel^{29,30}, Patrice Nordmann^{29,30},
7 Timothy G Palzkill^{9,10}, George A Jacoby²³, Karen Bush¹¹, Robert A Bonomo^{1-6,12,13,‡}.

8

9 Departments of ¹ Molecular Biology and Microbiology, ² Medicine, ³ Pharmacology, ⁴ Biochemistry, and
10 ⁵ Proteomics and Bioinformatics, Case Western Reserve University, Cleveland, OH, USA.

11 ⁶ Research Service, Louis Stokes Cleveland Department of Veterans Affairs, Cleveland, OH, USA.

12 ⁷ National Center for Biotechnology Information, National Library of Medicine, National Institutes of
13 Health, Bethesda, MD, 20894, USA.

14 ⁸ Antimicrobial Development Specialists, LLC, Nyack, NY, USA.

15 ⁹ Department of Pharmacology and Chemical Biology and ¹⁰ Verna Marrs McLean Department of
16 Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX, USA.

17 ¹¹ Department of Biology, Indiana University Bloomington, Bloomington, IN, USA.

18 ¹² CWRU-Cleveland VAMC Center for Antimicrobial Resistance and Epidemiology (Case VA CARES),
19 Cleveland, OH, USA.

20 ¹³ Geriatric Research Education and Clinical Centers (GRECC), Louis Stokes Cleveland Department of
21 Veterans Affairs, Cleveland, OH, USA.

- 22 ¹⁴ Instituto de Biología Molecular y Celular de Rosario (IBR, CONICET-UNR), Ocampo y Esmeralda,
23 Rosario, Argentina.
- 24 ¹⁵ Plataforma Argentina de Biología Estructural y Metabolómica PLABEM, Buenos Aires, Argentina.
- 25 ¹⁶ Plataforma Argentina de Biología Estructural y Metabolómica PLABEM, Buenos Aires, Argentina.
- 26 ¹⁷ Área Biofísica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario,
27 Rosario, Argentina.
- 28 ¹⁸ Centro de Investigaciones en Química Biológica de Córdoba (CIQUIBIC), CONICET, Universidad
29 Nacional de Córdoba, Córdoba, Argentina.
- 30 ¹⁹ Departamento de Química Biológica Ranwel Caputto, Facultad de Ciencias Químicas, Universidad
31 Nacional de Córdoba, Córdoba, Argentina.
- 32 ²⁰ Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy.
- 33 ²¹ Microbiology and Virology Unit, Florence Careggi University Hospital, Florence, Italy.
- 34 ²² Centre for Protein Engineering (CIP), University of Liège, Liège, Belgium.
- 35 ²³ Lahey Hospital and Medical Center, Burlington, Massachusetts, USA.
- 36 ²⁴ Red Española de Investigación en Patología Infecciosa (REIPI), Instituto de Salud Carlos III, Madrid,
37 Spain.
- 38 ²⁵ Servicio de Microbiología, Hospital Universitario Son Espases, Instituto de Investigación Sanitaria Illes
39 Balears (IdISBa), Palma de Mallorca, Spain.
- 40 ²⁶ Laboratoire de Bactériologie, Centre Hospitalier Régional Universitaire, Besançon, France.
- 41 ²⁷ UMR6249 CNRS Chrono-Environnement, Université de Bourgogne Franche-Comté, Besançon,
42 France.

43 ²⁸ Centre National de Référence de la résistance aux antibiotiques, Besançon, France.

44 ²⁹ Emerging Antibiotic Resistance Unit, Medical and Molecular Microbiology, Faculty of Science and
45 Medicine, University of Fribourg, Fribourg, Switzerland.

46 ³⁰ Swiss National Reference Center for Emerging Antibiotic Resistance (NARA), University of Fribourg,
47 Fribourg, Switzerland.

48 ³¹ UCL School of Pharmacy, University College London, London, United Kingdom.

49 ³² Department of Medical Biotechnology, University of Siena, Siena, Italy.

50 ³³ Department of Microbiology and Immunology, Creighton School of Medicine, Omaha, NE, USA.

51

52

53

54

55

56 Corresponding Author:

57 ‡ Robert A. Bonomo, MD, Medical Service, Louis Stokes Cleveland Department of Veterans Affairs

58 Medical Center, 10701 East Blvd., Cleveland, OH 44106. P: 216-791-3800, x 4645. Email:

59 robert.bonomo@va.gov.

60 **Abstract**

61 Unlike classes A and B, a standardized amino acid numbering scheme has not been proposed for
62 the class C (AmpC) β -lactamases, which complicates communication in the field. Here, we propose a
63 scheme developed through a collaborative approach that considers both sequence and structure, preserves
64 traditional numbering of catalytically important residues (Ser⁶⁴, Lys⁶⁷, Tyr¹⁵⁰, and Lys³¹⁵), is adaptable to
65 new variants or enzymes yet to be discovered, and includes a variation for genetic and epidemiological
66 applications.

67

68 An urgent need exists to address current inconsistencies in the numbering of amino acid residues
69 among class C β -lactamases, both within families and across the class. Established conventions in the
70 field define three common features shared among the serine-type β -lactamases. In the class C β -
71 lactamases, also known as AmpC β -lactamases, these features occur at recognizable conserved motifs:
72 S⁶⁴XXK (where S⁶⁴ is the active site serine); Y¹⁵⁰XN; and K³¹⁵(S/T)G (1–4). These designations align
73 with the amino acid sequence of the mature form of both the P99 AmpC (NCBI RefSeq Accession
74 WP_049134845.1 - originally characterized in an *Enterobacter cloacae* strain – now found to be an *E.*
75 *hormaechei* strain – in GenBank Accession CAA30257.1), and *Escherichia coli* AmpC (NCBI RefSeq
76 Accession WP_001336292.1 – originally characterized in strain K-12 found in GenBank Accession
77 AAC77110.1). While *E. coli* AmpC has historical significance as both the first β -lactamase reported (5)
78 and the first class C β -lactamase sequenced, P99 maintains identical numbering of conserved motifs while
79 the mature form begins with a natural residue one (6). Many other class C β lactamases, however, possess
80 insertions and deletions that shift the numbering of the conserved residues, significantly complicating
81 both nomenclature and comparisons between enzymes.

82 For this report, we analyzed 155 unique AmpC structures deposited in the Protein Data Bank
83 (including 142 supported by 66 publications), and found that 129 β -lactamase structures identify the
84 catalytic serine as Ser⁶⁴ (123 naturally and 6 with alignment), 10 number from the beginning of the
85 precursor form with the signal peptide included, and the remaining 16 number from the beginning of the
86 mature form, but do not identify the catalytic serine as Ser⁶⁴ (of which 8 are not associated with a
87 publication). Additionally, based on a literature search of PubMed, we found consistency is lacking for
88 numbering within the various families of class C β -lactamases. As an example, since the term PDC
89 (*Pseudomonas*-derived cephalosporinase) was coined in 2009 for the chromosomal AmpC of
90 *Pseudomonas aeruginosa*, three different approaches have been used to number amino acid residues in
91 this β -lactamase (7). These approaches include: *i*) direct numbering of residues beginning with the N-
92 terminus of the precursor protein (7); *ii*) direct numbering of residues beginning with the N-terminus of

93 the mature protein (8); and *iii*) alignment-based numbering designed to maintain the conventional
94 assignment of conserved residues and to simplify numbering for comparisons across families (9).
95 Unfortunately, it can be unclear to readers which of the various schemes is being used in a given
96 publication. As a result, authors may sometimes find choosing a numbering scheme and numerically
97 designating a given residue problematic. Comparing findings from multiple publications may be made
98 unnecessarily difficult; resolving ambiguity in assignment may be extremely challenging. For example, a
99 reference to Gly at position 183 in PDC could refer to a site that is described as having a clinically
100 relevant mutation if numbering begins with Met¹ of the precursor form, but would refer to a different
101 glycine, 26 residues away, using alignment based numbering (10, 11).

102 To address this growing concern, we propose a numbering scheme to use consistently when
103 referring to crystallographically equivalent positions in the mature form of any class C β -lactamase. We
104 suggest the acronym “*SANC*” to name the scheme, for Structural Alignment-based Numbering of class C
105 β -lactamases, or else the simpler term “structural position.” In developing this numbering scheme, we
106 adapted the approaches used by Ambler et al. for the class A β -lactamases (12) and Galleni et al. for the
107 class B β -lactamases (13). We conducted an amino acid alignment of 32 AmpC β -lactamases, both
108 chromosomal and plasmid encoded (**Supplemental Material**) and identified characteristic differences
109 from P99 for each enzyme (**Table 1** Sequences were obtained from the National Center for
110 Biotechnology Information Protein Database (14) and signal peptide cleavage sites were determined using
111 Uniprot (or SignalP 5.0 for entries not present in Uniprot) (15, 16). Mature protein sequences were
112 aligned using the MUSCLE algorithm (17) with default settings.

113 Consensus secondary structure (defined as a majority of structures in agreement for a given
114 amino acid position) was determined based on comparisons of a representative structure of each of the ten
115 AmpC β -lactamases for which one or more structures are available in the Protein Data Bank, specifically:
116 ACT-1 (PDB: 2ZC7), ADC-7 (PDB: 4U0T), CMY-2 (PDB: 1ZC2), *E. coli* AmpC (PDB: 2BLS), FOX-4
117 (PDB: 5CGS), MOX-1 (PDB: 3W8K), *Mycobacterium smegmatis* AmpC (PDB: 5E2H), PDC-1 (PDB:

118 4GZB), and TRU-1 (PDB: 6FM6). The consensus agrees with the secondary structure (or lack thereof) of
119 P99 for just over 90% of residues. This consensus was used to annotate secondary structure, including
120 stripes to indicate residues with an even split between two secondary structure types, and helix numbers
121 on the alignment. Finally, a simple literature survey was conducted to determine residues belonging in
122 either the consensus portion or fullest likely extent of the Ω -loop or R2-loop, both of which are also
123 annotated on the alignment. By including this structural information, we hope to both better correlate the
124 numbering system with well-known structural features and to provide additional points of reference for
125 those just beginning to work with AmpC structures.

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

130 Amino acid numbering was based on *E. cloacae* complex P99 while preserving the conventional
131 numbering of the following residues: Ser⁶⁴, Lys⁶⁷, Tyr¹⁵⁰, and Lys³¹⁵. Insertions relative to P99 were
132 addressed by appending lowercase letter(s) to the number of the amino acid immediately preceding the
133 insertion (e.g., 125a in PDC-1). Deletions relative to P99 were skipped, resulting in “ghost residues” (e.g.,
134 ACC-1 has residues G115 and L117 with a deleted residue at 116). For mature enzymes with more C-
135 terminal amino acid residues than P99, additional residues are assigned numbers in numerical order
136 beginning with 362. For mature enzymes with more N-terminal amino acid residues than P99, the first
137 additional residue is numbered 0 and subsequent residues are numbered by appending a lowercase letter
138 to zero while moving in an N-terminal direction (e.g., 0 and 0a for BUT-1 and *Edwardsiella* AmpC).
139 Signal peptide residues are assigned negative numbers, beginning with -1 for the residue adjacent to the
140 cleavage site and proceeding in the N-terminal direction until all residues are numbered. Multiple
141 sequence alignments are not considered for the signal peptide regions. **Figure 1** illustrates these principles
142 with several examples.

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

148 Providing numbering under both schemes is essential to our proposal. Structural numbering
149 maintains continuity with the conventional assignment of the catalytic serine as Ser⁶⁴ and the majority of
150 existing literature on class C β -lactamase structure and function while precursor numbering enables direct
151 gene translation and simplifies interpretation of sequencing results, particularly within a single family.
152 Utilizing this hybrid approach, an initial description of a typical PDC variant might read “PDC-221
153 differs from PDC-1 (GenBank AAG07497.1) by a single amino acid substitution, E247K, occurring at
154 *SANC* position 219.”

155 In the supplementary materials, we provide a table featuring a multiple sequence alignment of 32
156 class C β -lactamases with column headers indicating the appropriate number to be used at each position.
157 The spreadsheet also features a text-based alignment of the structures used in determining the consensus
158 secondary structure. Separately, we provide a protein profile hidden Markov model (HMM) which
159 implements the *SANC* scheme, built from the multiple sequence alignment using HMMER
160 (<http://hmmer.org>). Alignments of the HMM to class C β -lactamases can be expected to produce correct
161 *SANC* assignments when results of the search are examined. We suggest using the HMM, rather than
162 examinations by eye, to make position assignments under this scheme for novel AmpC enzymes that may
163 be discovered in the future. Finally, basic instructions for using our HMM with the HMMER software are
164 also included with the supplementary materials.

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

168 **Acknowledgements**

169 The authors would like to thank Ram Podicheti and Cameron Divoky for sequencing and species
170 verification of the P99 producing *Enterobacter hormaechei* strain originally provided by Mark Richmond.

171 **Funding**

172 Research reported in this publication was supported by the National Institute of Allergy and
173 Infectious Diseases of the National Institutes of Health (NIH) to R.A.B. under Award Numbers
174 R01AI100560, R01AI063517, and R01AI072219. This study was also supported in part by funds and/or
175 facilities provided by the Cleveland Department of Veterans Affairs to K.M.P.-W. and R.A.B., Veterans
176 Affairs Merit Review Program Award Number 1I01BX002872 to K.M.P.-W and Veterans Affairs Merit
177 Review Program Award Number 1I01BX001974 to R.A.B. from the Biomedical Laboratory Research &
178 Development Service of the VA Office of Research and Development, and the Geriatric Research
179 Education and Clinical Center VISN 10. N.D.H. receives funding from Alimatrix, Biomerieux, Merck,
180 Roche, and Shionogi and serves as a scientific advisor for Streck. The work of Michael Feldgarden,
181 Daniel H. Haft, and William Klimke was supported by the Intramural Research Program of the National
182 Library of Medicine, National Institutes of Health.

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

Class C β -Lactamase	NCBI Accession	Insertions and Deletions Relative to <i>E. cloacae</i> complex P99
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
<i>B. multivorans</i> 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
<i>E. coli</i> AmpC	WP_001336292.1	-1, -2, -3
EC-5	WP_001443153.1	-1, -2, -3
<i>Edwardsiella</i> 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
<i>M. smegmatis</i> 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

186 **Table 1:** Insertions and deletions present in the AmpC enzymes examined when compared to *E. cloacae*
187 complex P99. Minus indicates a deletion and plus indicates an insertion. Appended letters indicate an
188 insertion follows a given residue number.

189

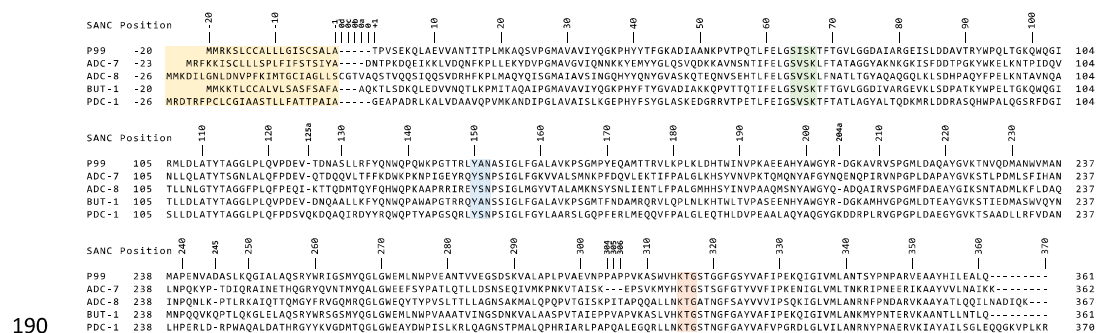


Figure 1: Using an alignment to assign SANC-based amino acid residue numbers. Positions corresponding to insertions and deletions are indicated in bold. ADC-7 adds residues 0, 204a, and 262 and deletes residues 245 and 304-306. ADC-8 adds residues 0-0d and 262-267 and deletes residue 245. BUT-1 adds residue 0. PDC-1 adds residues 125a, 204a, and 362-370 and deletes residue 245. For reference, signal sequences are highlighted in yellow, S⁶⁴XXK in green, Y¹⁵⁰XN in blue, and K³¹⁵(S/T)G in red.

197 **References**

- 198 1. Bush K. 2013. The ABCD's of β -lactamase nomenclature. *Journal of Infection and Chemotherapy*
199 19:549–559.
- 200 2. Jacoby GA. 2009. AmpC β -lactamases. *Clin Microbiol Rev* 22:161–182.
- 201 3. Oefner C, D'Arcy A, Daly JJ, Gubernator K, Charnas RL, Heinze I, Hubschwerlen C, Winkler FK.
202 1990. Refined crystal structure of β -lactamase from *Citrobacter freundii* indicates a mechanism for β -
203 lactam hydrolysis. *Nature* 343:284–288.
- 204 4. Lobkovsky E, Moews PC, Liu H, Zhao H, Frere JM, Knox JR. 1993. Evolution of an enzyme
205 activity: crystallographic structure at 2-Å resolution of cephalosporinase from the ampC gene of
206 *Enterobacter cloacae* P99 and comparison with a class A penicillinase. *PNAS* 90:11257–11261.
- 207 5. Abraham EP, Chain E. 1940. An enzyme from bacteria able to destroy penicillin. *Nature* 146:837.
- 208 6. Galleni M, Lindberg F, Normark S, Cole S, Honore N, Joris B, Frere JM. 1988. Sequence and
209 comparative analysis of three *Enterobacter cloacae* ampC β -lactamase genes and their products.
210 *Biochemical Journal* 250:753–760.
- 211 7. Rodríguez-Martínez J-M, Poirel L, Nordmann P. 2009. Extended-spectrum cephalosporinases in
212 *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 53:1766–1771.
- 213 8. Berrazeg M, Jeannot K, Enguéné VYN, Broutin I, Loeffert S, Fournier D, Plésiat P. 2015. Mutations
214 in β -lactamase AmpC increase resistance of *Pseudomonas aeruginosa* isolates to antipseudomonal
215 cephalosporins. *Antimicrob Agents Chemother* 59:6248–6255.

- 216 9. Drawz SM, Taracila M, Caselli E, Prati F, Bonomo RA. 2011. Exploring sequence requirements for
217 C₃/C₄ carboxylate recognition in the *Pseudomonas aeruginosa* cephalosporinase: Insights into
218 plasticity of the AmpC β -lactamase. *Protein Sci* 20:941–958.
- 219 10. MacVane SH, Pandey R, Steed LL, Kreiswirth BN, Chen L. 2017. Emergence of ceftolozane-
220 tazobactam-resistant *Pseudomonas aeruginosa* during treatment is mediated by a single AmpC
221 structural mutation. *Antimicrob Agents Chemother* 61:e01183-17.
- 222 11. Cabot G, Bruchmann S, Mulet X, Zamorano L, Moyà B, Juan C, Haussler S, Oliver A. 2014.
223 *Pseudomonas aeruginosa* ceftolozane-tazobactam resistance development requires multiple
224 Mutations leading to overexpression and structural modification of AmpC. *Antimicrob Agents*
225 *Chemother* 58:3091–3099.
- 226 12. Ambler RP, Coulson AF, Frère JM, Ghuyssen JM, Joris B, Forsman M, Levesque RC, Tiraby G,
227 Waley SG. 1991. A standard numbering scheme for the class A β -lactamases. *Biochem J* 276 (Pt
228 1):269–270.
- 229 13. Galleni M, Lamotte-Brasseur J, Rossolini GM, Spencer J, Dideberg O, Frère J-M, The Metallo- β -
230 Lactamase Working Group. 2001. Standard numbering scheme for class B β -lactamases.
231 *Antimicrobial Agents and Chemotherapy* 45:660–663.
- 232 14. NCBI Resource Coordinators. 2018. Database resources of the National Center for Biotechnology
233 Information. *Nucleic Acids Res* 46:D8–D13.
- 234 15. The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids*
235 *Research* 47:D506–D515.

- 236 16. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne
237 G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat*
238 *Biotechnol* 37:420–423.
- 239 17. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
240 *Nucleic Acids Res* 32:1792–1797.
- 241

SANC Position		-20		-10		10		20		30		40		50		60		70		80		90		100															
P99	-20	MMRKS	LCCALL	LGISCS	ALA----	TPVSE	QLA	EVVANT	ITPL	MKAQ	SVPG	MAVA	IYQG	KPHY	YTFG	KADIA	AANK	PVTP	QTLF	ELGS	ISKTF	TGVL	GGDA	IARGE	ISLDD	AVTRY	WPQL	TGKQ	WQGI	104									
ADC-7	-23	MRFKK	ISCLLL	SPLF	IFST	SIYA---	DNTPK	DQEI	IKLL	VQNF	KPLLE	KYDV	PGMA	VGIQ	NNKYE	MYGL	QSVQ	DKKAV	NSNT	IFEL	GSVSK	LFAT	AGGY	AKNG	KISF	DDTP	PGKY	WKLK	NTPI	DQV	104								
ADC-8	-26	MMKD	ILGNL	DNVPF	KIMT	GCIAG	LLS	CGTVA	QSTV	QSQI	QSVDR	HFPL	MAQY	ISGM	IAVSI	NGHY	YQNY	GVASK	QTEQ	NVSE	HTEF	ELGS	VSKL	FNAT	LTGY	AAQ	QQLK	LS	DP	APQY	FPEL	KNTA	VNQ	104					
BUT-1	-20	MMKK	LCCAL	LVLS	SASF	SAFA---	AQTL	LSDK	QLED	VVNQ	TLK	PMIT	AQA	IPGM	AVAV	IYQG	KPHY	YTYG	VAD	IAKQ	PVTT	QTIF	ELGS	VSKT	FTG	VL	GGD	I	V	ARG	EVK	LS	DP	ATKY	WPEL	TGKQ	WQGI	104	
PDC-1	-26	MRDR	FRPCL	CGIA	ASTLL	FATPA	IA---	GEAP	ADR	LKAL	VAAV	QPMK	AND	IPGL	AVAI	LSL	GEPHY	FSYGL	ASK	ED	GRR	VTP	ETL	FEI	GS	VSKT	FTAT	L	AG	YAL	TQ	KMR	LDD	RAS	QH	WPAL	QGS	RFDGI	104

SANC Position		110		120		125a		130		140		150		160		170		180		190		200		204a		210		220		230																																							
P99	105	RMLD	LATY	TAGG	LPLQ	VP	DEV-	TDNA	SLLR	FYQN	WQPK	PGT	TRLY	ANAS	IGL	FGAL	AVK	PSG	MPYE	QAMT	TRVL	KPLK	LDHT	INVP	KAE	E	AHY	ANG	YR-	DGK	AVR	VSP	GMLD	AQAY	GVK	TNV	QDM	ANV	VMAN	237																													
ADC-7	105	NLLQ	LATY	TSGN	LALQ	FP	DEV-	QTDQ	QVLT	FFK	DKW	KPN	PIGE	YRQ	YSNP	SIGL	FGV	VVAL	SMNK	PFQ	VLEK	TFP	ALG	LKHS	VVNV	PKT	QM	QNYA	FG	YN	QENQ	PIR	VNP	GPLD	APAY	GVK	STL	PD	MLS	FIHAN	237																												
ADC-8	105	TLLN	LGTY	TAGG	FLQ	PEQI-	KTTQ	DMT	QYF	QH	WQ	PKA	PRR	I	REYS	NPSI	GLM	GYV	TAL	MKNS	YSLI	ENTL	FP	ALG	MH	HSV	IN	V	PA	QMS	NY	ANG	YQ-	ADQ	AI	R	VSP	GM	FD	EA	YGI	KSN	TAD	M	KFL	DAQ	237																						
BUT-1	105	TLLD	LATY	TAGG	LPLQ	VP	DEV-	DNQA	ALLK	FYQN	WQ	PA	APG	TRR	QYAN	SSIGL	FGAL	AVK	PSG	M	T	F	D	AM	RQ	RVL	Q	PLN	L	K	HTL	TV	PA	SE	E	N	H	Y	ANG	YR-	DGK	AM	H	V	G	P	M	L	D	T	E	A	Y	G	V	K	S	T	I	E	D	M	A	S	H	V	Q	YN	237
PDC-1	105	SLLD	LATY	TAGG	LPLQ	FP	DEV-	DNQA	ALLK	FYQN	WQ	PA	APG	TRR	QYAN	SSIGL	FGAL	AVK	PSG	M	T	F	D	AM	RQ	RVL	Q	PLN	L	K	HTL	TV	PA	SE	E	N	H	Y	ANG	YR-	DGK	AM	H	V	G	P	M	L	D	T	E	A	Y	G	V	K	S	T	I	E	D	M	A	S	H	V	Q	YN	237

SANC Position		240		245		250		260		270		280		290		300		304		308		310		320		330		340		350		360		370																																																																																											
P99	238	MAPEN	VAD	ASL	KQGI	AL	AQ	SRY	WR	IG	SM	YQ	GL	GW	EMLN	NP	VE	ANT	V	VEG	S	D	S	K	V	A	L	A	P	L	V	A	E	V	N	P	P	A	P	P	K	A	S	W	V	H	K	T	G	S	T	G	F	G	S	Y	V	A	F	I	P	E	K	Q	I	G	I	V	M	L	A	N	T	S	Y	P	N	P	A	R	V	E	A	A	H	I	L	E	A	L	Q	-----	361																																
ADC-7	238	LNPQ	KYP-	TDI	Q	RAI	N	E	T	H	Q	G	R	Y	Q	N	T	M	Y	Q	A	L	G	W	E	F	S	P	A	T	L	Q	T	L	L	D	S	N	S	E	Q	I	V	M	K	N	K	V	T	A	I	S	K	---	E	P	S	V	K	M	Y	H	K	T	G	S	T	S	G	F	G	T	Y	V	V	F	I	P	K	E	N	I	G	L	V	M	L	T	N	K	R	I	P	N	E	E	R	I	K	A	A	Y	V	V	L	N	A	I	K	-----	362														
ADC-8	238	INPQ	N	L	K-	P	T	L	R	K	A	I	Q	T	T	M	G	Y	F	R	V	G	Q	M	R	Q	L	G	W	E	Y	T	P	V	S	L	T	T	L	A	G	N	S	A	K	M	A	L	Q	P	V	T	G	I	S	K	P	I	T	A	P	Q	A	L	N	K	T	G	A	T	N	G	F	S	A	Y	V	V	I	P	S	Q	I	G	L	V	M	L	A	N	R	N	F	P	N	D	A	R	V	A	A	Y	A	T	L	Q	I	L	N	A	D	I	Q	K	---	367									
BUT-1	238	MNPQ	Q	V	K	Q	P	T	L	Q	G	L	E	L	A	Q	S	R	Y	W	R	S	G	S	M	Y	Q	L	G	W	E	M	L	N	P	V	A	A	A	T	V	I	N	G	S	D	N	K	V	A	L	A	A	S	P	V	A	I	E	P	P	V	A	P	V	K	A	S	L	V	H	K	T	G	S	T	N	G	F	G	A	Y	V	A	F	I	P	E	K	Q	I	G	I	V	M	L	A	N	K	M	Y	P	N	T	E	R	V	K	A	A	N	T	L	N	T	L	Q	-----	361						
PDC-1	238	LHPER	L	D-	RP	WA	Q	A	L	D	A	T	H	R	G	Y	Y	K	V	G	M	T	Q	L	G	W	E	A	D	W	P	I	S	L	K	R	L	Q	A	G	N	S	T	P	M	A	L	Q	P	H	R	I	A	R	L	P	A	P	Q	A	L	E	G	Q	R	L	L	N	K	T	G	S	T	N	G	F	G	A	Y	V	A	F	V	P	G	R	D	L	G	L	V	I	L	A	N	R	N	Y	P	N	A	E	R	V	K	I	A	Y	A	I	L	S	G	L	E	Q	Q	K	V	P	L	K	R	-----	370