AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Complete Genome Sequence of *Escherichia coli* BLR(DE3), a *recA*-Deficient Derivative of *E. coli* BL21(DE3)

Philippe Goffin,[a,b] Philippe Dehottay[a]

GSK Vaccines, Rixensart, Belgium[a]; Laboratoire de Génétique et Physiologie Bactérienne, IBMM, Faculté des Sciences, Université Libre de Bruxelles (ULB), Gosselies, Belgium[b]

ABSTRACT *Escherichia coli* BLR(DE3) is a commercially available *recA*-deficient derivative of BL21(DE3), one of the most widely used strains for recombinant protein expression. Here, we present the full-genome sequence of BLR(DE3) and highlight additional differences with its parent strain BL21(DE3) which were previously unreported but may affect its physiology.

Received 10 April 2017 **Accepted** 17 April 2017 **Published** 1 June 2017

Citation Goffin P, Dehottay P. 2017. Complete genome sequence of *Escherichia coli* BLR(DE3), a *recA*-deficient derivative of *E. coli* BL21(DE3). Genome Announc 5:e00441-17. https://doi.org/10.1128/genomeA.00441-17.

Address correspondence to Philippe Goffin, philippe.x.goffin@gmail.com.

Escherichia coli is widely used for the production of recombinant polypeptides, including therapeutic proteins (1). The most popular system for protein overexpression in *E. coli* is the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible T7 RNA polymerase cascade system, in which expression of the gene of interest is controlled by a T7 RNA polymerase promoter (e.g., on a pET vector) in a strain carrying the T7 RNA polymerase gene under control of an IPTG-inducible promoter (e.g., λDE3 lysogenic strains) (2). The genome sequence of the widely used *E. coli* B strain BL21(DE3) (2) is available (3). BLR(DE3) is a *recA*-deficient derivative of BL21(DE3) that helps stabilize plasmids containing repetitive sequences (4). According to Novagen, BLR(DE3) has the same genotype, except for a mutation in *recA* [Δ(*srl-recA*)306::Tn*10*], which was initially obtained in an *E. coli* K-12 background (5) and later transferred to BL21 (4, 6).

We sequenced the BLR(DE3) strain obtained from Novagen in order to verify whether additional differences are present in BLR(DE3) compared to BL21(DE3). We performed *de novo* hybrid assembly from PacBio sequencing (102,708 reads, with a mean length of 5,173 bp) and Illumina HiSeq 2000 sequencing (approx. 4,500,000 high-quality paired-end 100-bp reads) for closing gaps and correcting PacBio sequencing errors. A total of 4,306 coding sequences (CDSs), 230 pseudogenes, 85 tRNA genes, 22 rRNA genes (8 operons), and 64 noncoding RNA genes were annotated.

A comparison with BL21(DE3) indicates larger differences than expected, with two main divergent regions. First, the DE3 prophage contains three large deletions of 4.9 kb (*cro-Rz*), 1.8 kb (*A-B*), and 11.1 kb (*Fi-J*), suggesting the prophage is nonfunctional. As expected, the second divergent region is the *recA* locus. A 6.9-kb region from *recA* to *srlR* is replaced with Tn*10* in BLR(DE3). Besides *recA* and the *srl* operon, BLR(DE3) thus also lacks *pncC* (encoding nicotinamide mononucleotide deamidase, a key enzyme of the pyridine nucleotide cycle [7]) and *mtlB* (encoding a lytic murein transglycosylase). Downstream of Tn*10*, a large region (approx. 75 kb; *srlR-queE*) diverges significantly from BL21(DE3), with 577 nucleotide substitutions (1 to 3 bp), short insertions/deletions (1 bp), and the replacement of an IS*186* element with a full clustered regularly interspaced short palindromic repeat (CRISPR) locus, as found in *E. coli* MG1655. We infer that the entire ~85-kb region from *recA* to *queE* was transferred from the original K-12 background bearing the Δ(*srl-recA*)306::Tn*10* mutation. This region contains several genes involved in metabolism and other key processes, such as DNA mismatch repair (*mutS*) or stress response (*rpoS*). Importantly, *rpoS* carries a nonsense mutation at codon 33, which has been shown to reduce σ^S activity (8). Outside the DE3 and *recA*

loci, BLR(DE3) and BL21(DE3) contain fewer than 20 differences. Of note, these include a previously identified mutation in *ilvA*, a gene responsible for isoleucine auxotrophy (9).

Globally, the genome sequence of BLR(DE3) indicates that it cannot be merely considered a *recA*-deficient BL21(DE3) strain. The newly identified differences might contribute to a differential physiology of BLR(DE3), besides its reported impaired homologous recombination.

Accession number(s). The complete genome sequence of *Escherichia coli* BLR(DE3) has been deposited at GenBank under accession number CP020368.

ACKNOWLEDGMENTS

This work was sponsored by GlaxoSmithKline Biologicals S.A., which provided the funding source, was involved in all stages of the study conduct and analysis, and took charge of the costs incurred in publishing.

We all are, or were at the time of study, employees of the GSK group of companies. P.D. reports ownership of shares in the GSK group of companies. P.G. and P.D. are named inventors on patents or patent applications related to the field of vaccine development and owned by GSK.

REFERENCES

1. Sanchez-Garcia L, Martín L, Mangues R, Ferrer-Miralles N, Vázquez E, Villaverde A. 2016. Recombinant pharmaceuticals from microbial cells: a 2015 update. Microb Cell Fact 15:33. https://doi.org/10.1186/s12934-016-0437-3.
2. Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol 189:113–130. https://doi.org/10.1016/0022-2836(86)90385-2.
3. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi SH, Couloux A, Lee SW, Yoon SH, Cattolico L, Hur CG, Park HS, Ségurens B, Kim SC, Oh TK, Lenski RE, Studier FW, Daegelen P, Kim JF. 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). J Mol Biol 394:644–652. https://doi.org/10.1016/j.jmb.2009.09.052.
4. Novagen. 2004. Competent cells user protocol. Novagen, Darmstatd, Germany.
5. Csonka LN, Clark AJ. 1979. Deletions generated by the transposon Tn*10* in the *srl recA* region of the *Escherichia coli* K-12 chromosome. Genetics 93:321–343.
6. Roca A. 1997. Initial characterization of mutants in a universally conserved RecA structural motif. PhD thesis. University of Wisconsin, Madison, Madison, WI.
7. Galeazzi L, Bocci P, Amici A, Brunetti L, Ruggieri S, Romine M, Reed S, Osterman AL, Rodionov DA, Sorci L, Raffaelli N. 2011. Identification of nicotinamide mononucleotide deamidase of the bacterial pyridine nucleotide cycle reveals a novel broadly conserved amidohydrolase family. J Biol Chem 286:40365–40375. https://doi.org/10.1074/jbc.M111.275818.
8. Subbarayan PR, Sarkar M. 2004. *Escherichia coli rpoS* gene has an internal secondary translation initiation region. Biochem Biophys Res Commun 313:294–299. https://doi.org/10.1016/j.bbrc.2003.11.132.
9. Schmidt M, Römer L, Strehle M, Scheibel T. 2007. Conquering isoleucine auxotrophy of *Escherichia coli* BLR(DE3) to recombinantly produce spider silk proteins in minimal media. Biotechnol Lett 29:1741–1744. https://doi.org/10.1007/s10529-007-9461-z.