Clinical Microbiology and Infection 23 (2017) 167–172

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Consensus statement

Consensus numbering system for the rifampicin resistance-associated *rpoB* gene mutations in pathogenic mycobacteria

E. Andre [1, 2, 3, *], L. Goeminne [1], A. Cabibbe [4], P. Beckert [5, 6], B. Kabamba Mukadi [1, 2], V. Mathys [7], S. Gagneux [8, 9], S. Niemann [6, 10], J. Van Ingen [2, 11], E. Cambau [3, 11, 12]

[1] *Pôle de Microbiologie Médicale, Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Brussels, Belgium*
[2] *Microbiology Unit, Cliniques Universitaires Saint-Luc, Brussels, Belgium*
[3] *European Society for Clinical Microbiology and Infectious Diseases (ESCMID) Study Group for Mycobacterial Infections (ESGMYC), ESCMID, Basel, Switzerland*
[4] *Emerging Bacterial Pathogens Unit, Tuberculosis Supranational Reference Laboratory (SRL), IRCCS San Raffaele Scientific Institute, Milano, Italy*
[5] *Molecular and Experimental Mycobacteriology Group, Research Center Borstel, Borstel, Germany*
[6] *German Center for Infection Research, Borstel Site, Borstel, Germany*
[7] *National Reference Laboratory for Tuberculosis and Mycobacterial Diseases, Scientific Institute of Public Health (WIV-ISP), Brussels, Belgium*
[8] *Swiss Tropical & Public Health Institute, Basel, Switzerland*
[9] *University of Basel, Basel, Switzerland*
[10] *Department of Medical Microbiology, Radboud University Medical Center, Nijmegen, The Netherlands*
[11] *APHP, Hôpital Lariboisière, Laboratory of Bacteriology, National Reference Center for Mycobacteria and Antimycobacterial Resistance, Paris, France*
[12] *University Paris Diderot, INSERM IAME UMR1137, Sorbonne Paris Cité, Paris, France*

ARTICLE INFO

Article history:
Received 29 July 2016
Received in revised form
12 September 2016
Accepted 12 September 2016
Available online 21 September 2016

Editor: F. Allerberger

Keywords:
Mycobacterium kansasii
Mycobacterium leprae
Mycobacterium tuberculosis
rpoB
Numbering system
Resistance
Rifampicin
Rifampin

ABSTRACT

The *rpoB* gene codes for the RNA polymerase β subunit, which is the target of rifampicin, an essential drug in the treatment of tuberculosis and other mycobacterial infections. This gene is present in all bacteria, but its length and nucleotide sequence vary between bacterial species, including mycobacteria. Mutations in the *rpoB* gene alter the structure of this protein and cause drug resistance. To describe the resistance-associated mutations, the scientific and medical communities have been using, since 1993, a numbering system based on the *Escherichia coli* sequence annotation. Using *E. coli* reference for describing mutations in mycobacteria leads to misunderstandings, particularly with the increasing use of whole genome sequencing, which brought an alternative numbering system based on the *Mycobacterium tuberculosis rpoB* sequence. We propose using a consensus numbering system for the reporting of resistance mutations based on the reference genomes from the species interrogated (such as strain H37Rv for *M. tuberculosis*). This manuscript provides the necessary figures and tables allowing researchers, microbiologists and clinicians to easily convert other annotation systems into one common language. **E. Andre, CMI 2017;23:167**

Introduction

Rifampicin, also named rifampin in the USA, is the key drug of the first-line treatment regimen for tuberculosis (TB), a disease affecting 9.6 million persons per year [1]. Multidrug-resistant TB (MDR-TB), defined as disease caused by *Mycobacterium tuberculosis* complex strains resistant to rifampicin and isoniazid, requires prolonged and more complex administration of alternative treatment regimens including second-line anti-TB drugs, and is associated with poorer treatment outcome [2,3]. Isoniazid-resistance is caused by different mutations affecting different genes [4], but the molecular basis of RIF resistance is simpler, as virtually all resistant strains present a mutation in the *rpoB* gene, and 95% of these are located within a small 81-bp region named the rifampicin-resistance determining region (RRDR) [5]. Consequently, molecular assays were easily developed to detect rifampicin resistance, and several commercial kits are now available and used worldwide [6–8]. Furthermore, molecular detection of rifampicin resistance is used as a surrogate

* Corresponding author.
E-mail addresses: emmanuel.andre@uclouvain.be, emmanuel.andre@outlook.be (E. Andre).

http://dx.doi.org/10.1016/j.cmi.2016.09.006

marker of MDR-TB in many countries, since <10% of rifampicin-resistant cases are not resistant to isoniazid [9–11]. The different species of the *M. tuberculosis* complex present similar resistance mutations [12–16].

Mutations in the *rpoB* gene are also associated with rifampicin resistance among other mycobacterial species, such as in *Mycobacterium leprae* causing leprosy [17,18] and *Mycobacterium kansasii* causing mainly respiratory infections in immunocompetent patients with also disseminated infections in immunocompromised patients [19]. Rifampicin resistance in *M. leprae* was described as associated with dapsone resistance in patients with MDR-leprosy. For these leprosy cases, the standard treatment regimen needs to be switched to a 2-year treatment combining daily second-line agents such as fluoroquinolones [20]. In *M. kansasii*, rifampicin is also a key component of treatment. Here *rpoB* gene mutations also lead to rifampicin resistance [19] and the need for alternative regimens [21].

Rifampicin resistance is caused by a structural alteration in the RNA polymerase β subunit, an enzyme coded by the *rpoB* gene. This mechanism was first described in 1981 for *Escherichia coli* [22,23], and it is therefore not specific to mycobacteria, although the clinical utility of *rpoB* mutations characterization is more important for TB and leprosy than for any other bacterial infections [24].

It was only in 1993 that Telenti *et al*. first suggested that rifampicin resistance in *M. tuberculosis* complex was associated with mutations in the *rpoB* gene, and observed a high similarity of the amino acids coded between several bacterial species including *E. coli*, *M. tuberculosis* complex and *M. leprae* [5]. Since the mutations described in that work were identified in an *rpoB* region that aligned with the RRDR of *E. coli*, Telenti *et al*. proposed an annotation system based on the *E. coli* genetic sequence. The authors probably underestimated how the diversity of the *rpoB* nucleotide sequence would later affect our ability to correctly report single nucleotide mutations causing amino acid substitutions [25–27]. Furthermore, although the authors recognized that resistance mutations may be located outside the RRDR, they did not pay attention to the low similarity between *E. coli* and *M. tuberculosis* complex in *rpoB* outside the RRDR. The existence and epidemiological importance of these 'non-RRDR' resistance mutations was later confirmed [28–30]. Nevertheless, this annotation system was rapidly adopted [31] and is still often used today for *M. tuberculosis* complex [32,33], *M. leprae* [34,35] and *M. kansasii* [19,36]. It was only recently that authors started using a species-specific numbering system [29,37–39]. This has caused some confusion because the same mutations are now named differently depending on the author, the year and the journal that published the work. For *M. leprae*, *rpoB* mutations were numbered first according to the gene sequence of *M. leprae rpoB* gene cloned in a cosmid, where the codon 425 corresponds to the codon 531 in *E. coli* and this was used to establish the WHO resistance surveillance network [40–42]. More recently, this expert group decided to switch to the numbering system of the *M. leprae* genome taking as a reference the genome sequence of the Tamil Nadu (TN) strain [37,43,44].

In this manuscript, we discuss the limitations of the traditional numbering system and propose to shift towards a more natural numbering system based on the mycobacterial reference sequences. This should allow better communication between laboratories, especially at a time when whole genome sequencing is becoming the standard for detecting drug resistance [39], improved disease surveillance [45,46] and faster spreading of scientific knowledge in the medical community.

Diversity of *rpoB* numbering systems

Among *M. tuberculosis* complex, *M. leprae* and *M. kansasii*, the nucleotide similarity lies between 86% and 88% for the entire *rpoB* gene and is slightly higher in the 81-bp RRDR (between 88.9% and 92.6%). The similarity of these mycobacteria with *E. coli* is less (47% and 74% for the *rpoB* gene and the RRDR respectively) [47].

Furthermore, the *rpoB* genes of *E. coli, M. tuberculosis* complex, *M. leprae* and *M. kansasii* differ in length (4029, 3519, 3537 and 3540 base pairs, respectively). These differences explain the variable shift between the *E. coli*-based annotation and the mycobacterial sequence-based numbering systems. At present, resistance mutations outside the RRDR have only been characterized for *M. tuberculosis* complex. This situation leads to an even more complex situation, as the shift from the *E. coli* numbering system to the *M. tuberculosis* complex numbering system is different depending on the location of the resistance mutation (i.e. +24 for the *M. tuberculosis* complex codon 170, and −81 for the other codons of interest).

Table 1 and Fig. 1 represent the codons that are most frequently associated with resistance mutations across the three mycobacterial species. *Mycobacterium tuberculosis* complex and *E. coli* present identical amino-acids only for eight of the nine codons, and not nine as initially suggested by Telenti *et al*. [5]. Only three of the nine codons show the same sequence.

Reporting systems derived from the use of commercial assays

The commercial assays have developed several mutation detection systems as an alternative to gene sequencing. These assays report rifampicin resistance based on the detection of mutations in the wild-type (WT) sequence of the RRDR by molecular beacons [48,49] or hybridization [8,37,50]. These tests include probes that bind to WT sequences (Xpert MTB/RIF (*M. tuberculosis*/rifampicin), Cepheid), eventually combined with additional probes binding to mutated sites (line probe assays InnoLipa RifTB, Innogenetics, Ghent, Belgium; GenoType MTBDRplus and GenoType lepraeDR, Hain Lifescience, Nehren, Germany), so silent mutations or mutations not associated with resistance can occasionally generate false-positive results [50–54].

The reporting format of these commercial assays can be 'matched' with the specific sequence covered by each probe (Xpert MTB/RIF), or the combination of WT and mutation bands (MTBDR*plus*, LepraeDR) as illustrated in Fig. 2.

In practice, the majority of *rpoB* mutations are located at *M. tuberculosis* complex codons 435, 445 and 450 (also named 516, 526 and 531, respectively, using the *E. coli* numbering system) and at codons 456 and 441 in *M. leprae*. As an example, in the presence of the mutation Ser450Leu, the Xpert MTB/RIF assay will report a negative 'Probe E', signing the presence of a mutation located between positions 447 and 452. For the same mutation, the MTBDR*plus* V2.0 assay will report a negative 'WT8' band associated with a positive 'MUT3' band. Although complex laboratory networks generally use a combination of these tests, the concordance between the results of each test is rarely verified in routine conditions because of the difficulties in comparing between the numbering systems. The number of commercial assays available for TB diagnostics and detection of rifampicin resistance will increase in the future [55], and each of these tests will probably come with its own reporting format. If this is the case, the efforts required for understanding and solving discordant results will increase further [32].

Consequences of different coexisting numbering systems

The sequencing methods used to identify mutations are currently shifting from traditional Sanger sequencing to whole genome sequencing [56,57]. Although both methods are used today and will continue to coexist for some years, high-throughput sequencing technologies present multiple advantages including