APPLICATION OF TOTAL ERROR APPROACH IN VALIDATION OF A BIOLOGICAL METHOD (ELISA) TO DETECT RESIDUES OF NICARBAZINE IN EGGS

Gaudin V, Laurentie M

Journal of Chromatography B

Contexte

La nicarbazine est un médicament vétérinaire de la famille des anticoccidiens. Ce principe actif est utilisé chez la volaille pour lutter contre les coccidioses, en supplémentation dans les aliments. Son utilisation est interdite chez les poules pondeuses. Toutefois, une contamination croisée des aliments pour la volaille lors de leur production peut entrainer la présence de résidus de nicarbazine dans les œufs. La nicarbazine possède un Niveau d'Action Différentiel (NAD) de 100 µg/kg. Nous avons donc utilisé cette limite comme la cible à atteindre par la méthode validée ici. La nicarbazine est un mélange équimolaire de 4,4-dinitrocarbanilide (DNC) et de 2-hydroxy-4,6 diméthyl pyrimidine (HDP). Le résidu marqueur est le DNC. Donc, comme dans la plupart des méthodes analytiques publiées, c'est le DNC qui était recherché dans les œufs, par le test ELISA sélectionné (CER, Belgique).

Méthodologie/Principaux résultats

Nous avons tout d'abord validé cette méthode selon la décision européenne 2002/657/CE [13]. Nous avons déterminé les caractéristiques de performance de la méthode selon cette décision, c'est-à-dire selon une approche « critère par critère ». L'objet de cet article est de montrer qu'à partir des résultats obtenus lors de la validation, nous pouvons aussi analyser les mêmes données grâce à une approche globale, c'est-à-dire selon une approche basée sur l'erreur totale et le profil d'exactitude. Nous pourrons ainsi conclure sur les performances de la méthode, en utilisant une autre approche et comparer les conclusions des deux approches quant aux performances de la méthode.

Dans un premier temps, pour construire le profil d'exactitude, nous avons analysé 6 standards d'étalonnage en tampon fournis avec le kit (0,2 ; 0,5 ; 2 ; 5 ; 10 et 20 µg/L). Six séries d'analyse ont été réalisées et chaque standard d'étalonnage a été analysé en double. Le nombre total de standards d'étalonnage était de 72, ce qui est suffisant pour établir la fonction de réponse. En effet, pour construire un modèle d'étalonnage logistique à 4 paramètres, il faut un minimum de 5 niveaux de concentration. De plus, 3 standards de validation (70, 140 et 210 µg/kg) dans les œufs ont été analysés. Après l'étape d'extraction des œufs, les extraits étaient dilués au 1/10 pour obtenir des niveaux de concentration rentrant dans le domaine couvert par la gamme d'étalonnage. Six séries d'analyse ont été réalisées et chaque standard de validation a été analysé en triple. Un total de 54 standards de validation a permis d'établir le profil d'exactitude. Nous avons alors procédé à une analyse quantitative des résultats.

Dans un deuxième temps, nous avons analysé 52 échantillons d'œufs blancs (sans résidus de nicarbazine) et 40 échantillons d'œufs supplémentés à 70 µg/kg en DNC. Nous avons alors effectué une analyse qualitative des résultats. Pour cela, nous avons fixé une valeur seuil, en utilisant la valeur du percentile 95, c'est-à-dire la valeur pour laquelle 95 % des échantillons sont en dessous. Au-dessus de cette valeur seuil, les échantillons sont considérés négatifs et inversement.

Tableau 21. Indices correspondant aux différents modèles de régression testés, classés selon l'indice d'Exactitude (I_A).

Modèle	I_A	I_{DR}	I_P	I_T
Régression Quadratique Pondérée ($1/X^2$)	0.6389	0.7183	0.7032	0.5162
Régression Quadratique Pondérée (1/X)	0.5614	0.4766	0.6085	0.6102
Régression Quadratique non Pondérée	0.5021	0.3403	0.5717	0.6506
Régression Logistique Pondérée à 4 Paramètres	0	0	0	0
Régression Logistique non Pondérée à 4 Paramètres	0	0	0	0
Régression Logistique Pondérée à 5 Paramètres	0	0	0	0
Régression Log-Log non Pondérée	0	0	0	0
Régression Logistique non Pondérée à 5 Paramètres	NC	NC	NC	NC
Régression Linéaire Pondérée	NC	NC	NC	NC
Régression Linéaire	NC	NC	NC	NC

I_A: indice d'Exactitude; I_{DR}: indice d'intervalle de dosage; I_T: indice de Justesse; I_P: indice de fidélité; NC: Indices impossibles à calculer

Nous avons analysé les résultats de validation grâce au logiciel e.noval, version 2.0. Nous avons utilisé différents indices qui reflètent les caractéristiques majeures de validation, pour sélectionner le meilleur modèle de fonction de réponse. Ces indices ont été définis dans un article de Rozet *et al.* [318]. Le **Tableau 19** résume les différents indices obtenus pour les différents modèles testés. En ce qui concerne l'indice d'exactitude I_A, qui est un indicateur global de la performance de la méthode, les meilleur indices sont obtenus pour les régressions quadratiques avec ou sans pondération. Toutefois, les valeurs d'indices sont basses et reflètent une grande variabilité et un manque de justesse. La meilleure valeur de justesse (I_T), parmi les 3 modèles précédemment cités, a été obtenue avec le modèle quadratique non pondéré, alors que les meilleurs indices d'exactitude et de fidélité ont été obtenus avec le modèle quadratique pondéré ($1/X^2$). Ce résultat montre que l'indice d'exactitude I_A seul ne peut servir au choix du modèle le mieux adapté. Dans le meilleur des cas (modèle quadratique pondéré ($1/X^2$)), seul 70% de l'intervalle de dosage est valide.

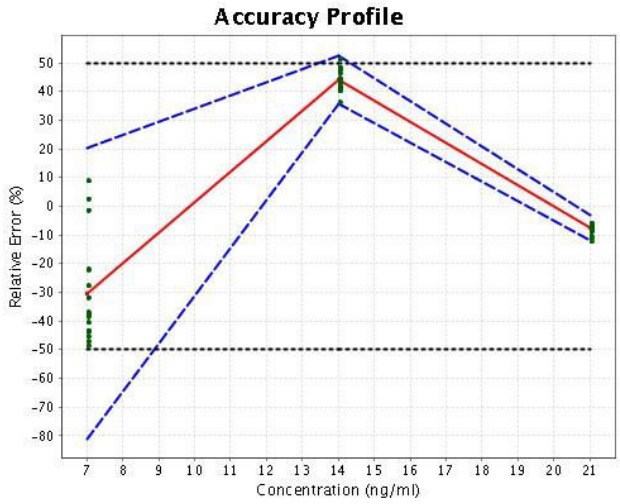
L'indice de fidélité I_P le meilleur a été obtenu pour le modèle quadratique pondéré ($1/X^2$). Cependant, la variabilité pour ce modèle à la concentration la plus basse est très élevée (23,62 % en fidélité intermédiaire). Pour les 3 modèles retenus plus haut, la répétabilité est conforme aux critères de la décision 2002/657/CE [13], mais pas la fidélité intermédiaire pour cette concentration. La répétabilité et la fidélité intermédiaire sont conformes aux critères réglementaires aux concentrations plus élevées. Le manque de justesse vient probablement d'un effet matrice. Pour résoudre ce problème de justesse et de fidélité, trois solutions ont été testées. Premièrement, les valeurs de concentrations ont été corrigées par un coefficient de correction. Le modèle corrigé est présenté dans la **Figure 47**. Dans ce cas, l'indice de justesse était de 1 mais la fidélité était moins bonne à tous les niveaux. La seconde solution était de travailler dans un intervalle de dosage plus restreint, de 14 à 21 µg/kg sans correction. Dans ce cas, le meilleur modèle était la régression quadratique pondérée (1/X) (**Figure 46)**.

La troisième solution était d'utiliser la méthode seulement comme une méthode qualitative, grâce à la valeur seuil déterminée à 20 µg/kg, ce qui est acceptable en regard du Niveau d'Action Différentiel (NAD) de 100 µg/kg. Cette valeur seuil correspond à un taux de faux-négatif inférieur ou égal à 5 %, comme exigé dans la réglementation européenne. Le taux de faux-positifs à cette valeur est de moins de 1 %.

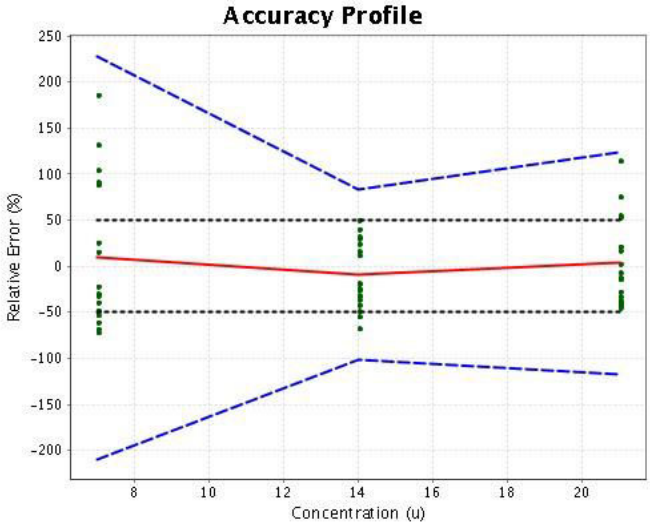
Il est impossible de comparer nos résultats issus de l'approche globale avec ceux de la validation de ce même kit réalisée par Huet et al. selon une approche critère par critère [319]. En effet, la méthode a été validée uniquement comme une méthode qualitative. Les auteurs avaient rapporté une capacité de détection CCβ inférieur à 3 µg/kg dans les œufs.

Figure 47. Profils d'exactitude obtenus à partir de la mesure de la concentration de nicarbazine dans les œufs (a) avec une régression quadratique pondérée (1/X) sans correction et (b) le modèle corrigé. Les lignes pleines rouges sont les biais relatifs, les lignes pointillées noires sont les limites de tolérance (intervalles de tolérance ß), les courbes bleues avec des tirets représentent les limites d'acceptabilité, les points représentent the concentrations relatives calculées des standards de validation.

(a) Modèle quadratique pondéré (1/X)



(b) Modèle corrigé



Conclusions/Importance

Nous avons pu analyser les données de validation du kit ELISA pour le dépistage de la nicarbazine dans les œufs, grâce à l'approche basée sur l'erreur totale et le profil d'exactitude. Les conclusions de l'approche globale sont que cette méthode ne peut être utilisée comme une méthode quantitative, en raison de son manque de justesse et de fidélité, dans l'intervalle de dosage qui nous intéressait. Cette méthode doit donc être utilisée seulement comme une méthode qualitative, ce qui est suffisant pour une méthode de dépistage. Cette étude a permis de conclure sur les performances de la méthode, en utilisant une nouvelle approche d'exploitation des données.

Quand nous avions validé ce kit selon la décision européenne 2002/657/CE, l'analyse qualitative des données avait montré une différence significative entre les œufs blancs et les échantillons supplémentés à 0.7*NAD. La capacité de détection était donc bien inférieure au NAD pour les œufs. Aucun résultat faux-positif, ni aucun résultat faux-négatif au niveau du NAD n'avait été obtenu. Donc les résultats étaient aussi satisfaisants au niveau qualitatif. Le kit avait montré une grande spécificité pour la nicarbazine. Il ne reconnaissait pas les grandes familles d'antibiotiques et reconnaissait très faiblement d'autres coccidiostatiques (réactions croisées 0.008 % à 0.025 %). Concernant l'analyse quantitative, les coefficients de variation (CV) intra-jour étaient compris entre 11.2 et 27.6% et les CV inter-jours entre 25.5 et 33.9 %. Donc, les conclusions de cette approche étaient similaires à celle de l'approche globale, c'est-à-dire que la méthode ne devrait pas être utilisée comme une méthode quantitative, étant donné les problèmes de répétabilité et surtout de fidélité intermédiaire. De plus, nous avions aussi conclu que la justesse n'était pas satisfaisante (comprise entre 48 et 67%).

Donc, l'approche critère par critère et l'approche globale du profil d'exactitude conduisent à la même conclusion, c'est-à-dire qu'il faut utiliser ce kit ELISA uniquement pour le dépistage qualitatif de la nicarbazine dans les œufs et non pour des analyses quantitatives. L'avantage du profil d'exactitude est de pouvoir comparer plusieurs modèles rapidement grâce au logiciel d'analyse des données et de visualiser graphiquement les problèmes de justesse et de fidélité, en fonction des concentrations. De plus, même si cela n'a pas fonctionné dans ce cas, il est possible grâce à un facteur de correction de corriger des effets matrices pour améliorer la justesse d'une méthode.

Journal of Chromatography B, 877 (2009) 2358–2362

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Application of total error approach to assess the performance of a biological method (ELISA) to detect nicarbazin residues in eggs☆

V. Gaudin[a], M. Laurentie[b,*]

[a] Veterinary Drug Residues Unit, Laboratory for the Research and Investigation of Veterinary and Disinfectants, Afssa Fougères, BP 90203, La Haute Marche, 35302 Fougères, France
[b] Pharmacokinetics-Pharmacodynamics Unit, Laboratory for the Research and Investigation of Veterinary and Disinfectants, Afssa Fougères, BP 90203, La Haute Marche, 35302 Fougères, France

ARTICLE INFO

Article history:
Received 29 August 2008
Accepted 3 March 2009
Available online 19 March 2009

Keywords:
Elisa
Eggs
Nicarbazin
Accuracy profile
Validation

ABSTRACT

Nicarbazin, a coccidiostat, is used as a feed additive in poultry but not in laying hens. Feed contamination may however occur resulting in residues being present in eggs. As a Maximum Residue Limit (MRL) does not exist for nicarbazin residues in eggs a "Differential Action Level" (DAL) of 100 μg/kg has been established by the Veterinary Medicines Directorate (VMD). We have studied a commercial ELISA kit validated to detect and quantify nicarbazin in eggs with a sensitivity of 3 μg/kg. We used the total error approach to assess the performance of and validate the kit at the DAL level. The accuracy profile has been successfully obtained for the ELISA kit. The method cannot however be validated as a semi-quantitative method and we have consequently determined a cut-off based on 5% false negative rate according to European Decision 2002/657 on blank and spiked samples (70 μg/kg). The cut-off value established was 20 μg/kg using the 95th percentile.

1. Introduction

Nicarbazin belongs to the family of coccidiostats and is used as a zootechnical feed additive for poultry. Coccidiostats are widely used compounds to prevent and treat coccidiosis, a contagious parasitic disease affecting livestock, particularly poultry, that is associated with warm and humid conditions [1].

According to Regulation 1831/2003/EC [2], anticoccidials are licensed as feed additives. Nicarbazin is authorised for use in broilers but not in laying hens [3]. Accidental cross-contamination of feed, however, has been shown to result in residues of the compounds in eggs [4,5]. Relatively high nicarbazin residue levels have also been found in the liver of poultry and a clear cause–effect relationship has been established between contaminated feed supplied from the feed mill and contaminated feed on the farm. Birds can also be exposed to alternative sources of nicarbazin near to slaughter including older nicarbazin-medicated feed from the feeding system or from litter. According to the Veterinary Medicines Directorate (VMD, United Kingdom), the likely cause of these residues is contamination at the feed mill, during transport and/or inadequate cleaning of hoppers and lines between batches of feed in farms [6].

Nicarbazin is an equimolar mixture of 4,4′-dinitrocarbanilide (DNC) and 2-hydroxy-4,6 dimethylpyrimidine (HDP). Its marker residue is DNC. Most of the residue analyses for nicarbazin are based on methods detecting the DNC molecule.

The European Commission has concerns over consumer health and has set Maximum Residue Limits (MRLs) for number of veterinary drugs in different matrices (Regulation 2377/90/EC) [7]. There is, however, no MRL for nicarbazin residues in eggs in Europe. In the absence of an MRL a zero tolerance approach should be used.

Several immunoassays for nicarbazin, using different platform technologies as ELISA, Biacore™, lateral flow device or dry chemistry immunoassay have been developed [8–11]. In particular, the developer of the ELISA kit (CER, Laboratory of Hormonology, Marloie, Belgium) has published the development and validation of this ELISA kit for nicarbazin in eggs according to decision EC/2002/657 [12]. The detection capability (Cc β), i.e. the smallest amount of the substance that can be detected, identified and/or quantified in a sample with an error probability of β, for the egg nicarbazin kit has been established as being 3 μg/kg [8].

In practice, some European Member States have used an alternative approach. In 1998, the VMD in the United Kingdom set a "Differential Action Level" (DAL) of 100 μg/kg body weight as a decision threshold for follow-up action.

Because the kit has only been validated in house by the producer, before it can be used in routine analysis by National Reference Laboratories (NRL), we wished to validate it in our laboratory at a DAL level of 100 μg/kg to verify its performance.

☆ This paper is part of a special issue entitled "Method Validation, Comparison and Transfer", guest edited by Serge Rudaz and Philippe Hubert.

* Corresponding author.
E-mail address: m.laurentie@fougeres.afssa.fr (M. Laurentie).

1570-0232/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2009.03.021

We have analysed the validation data using the total error approach in this work. The total error approach is based on constructing an accuracy profile [13–15], and has been widely discussed for physicochemical methods as described in this special review and in a previously published work [16]. In 2003, the American Association of Pharmaceutical Scientist (AAPS) and the Food and Drug Administration (FDA) [17] recommended that this approach be used for macromolecule ligand binding assays.

We therefore proposed to use this approach to assess the performance of an ELISA method in this study.

To construct an accuracy profile a validation study determining trueness of fit and precision simultaneously is required. From the results of the validation experiments a two-sided β expectation tolerance interval was calculated for each concentration level and the accuracy profile was constructed. We needed to define acceptance limits and the risk of the procedure resulting in a $(1-\beta)$ proportion of measurements falling outside of these limits. The risk was set at 5% and acceptance limits at ±50% in accordance with European Decision 2002/657 [12] and as described previously [16].

2. Materials and methods

2.1. immunoassay procedure

We used the CER ELISA kit (ref E.E.2, Laboratory of Hormonology, Marloie, Belgium).

The test principle and sample extraction methods have been extensively described by Huet et al. [8].

We used different apparatus to that described by Huet et al: a micro-titre plate washer ELP40 (ADIL Instruments, France) and micro-titre plate reader SpectraCount™ (Packard, France) to wash and read the micro-titre plate.

2.2. Total error profile

2.2.1. Calibration standards

Calibration standards were contained ready-to-use in the kit box from the manufacturer. Vials containing standard solutions of DNC of 0.2, 0.5, 2, 5, 10 and 20 μg/l in buffers were provided.

Six series of analysis were performed and calibration samples were analysed in duplicate. The total number of calibration standards are 72 and are sufficient to establish the response function including 4p or 5p logistic models as recommended by Hubert et al. [14].

2.2.2. Validation standards

Validation standards were prepared at 70, 140 and 210 μg/kg in homogenised eggs and were diluted after extraction (1/10), to obtain similar levels to those of the calibration standards before loading onto the micro-titre plate.

Six series of analyses were performed for the validation standards and were measured in triplicate. The total number of validation standards are 54 and are sufficient to establish the accuracy profile as recommended by Hubert et al. [14].

2.2.3. Validation analysis

The validation data were processed on e.noval software, version 2.0, and Seelva Version 1.0 beta 8, for logistic functions (Arlenda, Liège, Belgium).

2.3. Cut-off assessment

In a second process we used the kit as a qualitative tool. To do this we set cut-off criteria and determined concentration values based on positive or negative responses by testing 52 blank samples and 40 samples spiked to 70 μg/kg. We used the 95th percentile value to establish the cut-off, i.e. the value at which samples are deemed to be positive.

3. Results and discussion

3.1. Response function and accuracy profile

The relationships between response and concentrations were analysed using different regression models: linear, weighted linear, quadratic, weighted quadratic, 4 or 5 parameter logistic functions, weighted logistic functions and log–log regression.

We used a number of indexes reflecting the major validation criteria to select the best model. These were defined in a recent paper by Rozet et al. [18] and are contained in the software used. The first is the accuracy index (I_A). This is a global indicator of method performance depending on dosing range index (I_{DR}), trueness index (I_T) and precision index (I_P). The I_{DR} indicates that fraction of the range which is valid; when $I_{DR} = 1$, the whole range studied is accepted. The trueness index is an index describing method bias. An index close to 1 implies that the method is almost unbiased. The precision index describes random variation. An index close to 1 indicates that the method offers good precision.

Table 1 summarises the different indices obtained for the response functions tested. The best I_A was obtained for quadratic regressions with or without weighting. The index values, however, were low, ranging from 0.63 to 0.50. Considering the I_{DR}, the best model was the weighted $(1/X^2)$ quadratic regression with a value of 0.72, i.e. only 70% of the dosing range is acceptable.

Several models have an I_A value of 0 indicating that trueness or precision are poor, outside of the acceptance limits and consequently these models cannot be used directly.

Accuracy profiles for seven response functions tested are shown in Fig. 1. Visual examination confirms considerable variation and lack of trueness.

No calculation was possible for three models (unweighted five parameter logistic regression, weighted linear regression, and linear regression).

3.2. Trueness

Examination of I_T values (Table 1) shows that the best value (unweighted quadratic regression) obtained from a different model to that which had the best I_A (weighted $(1/X^2)$ quadratic regression). This indicates that the choice of model must be based not only on I_A but must take all indices into account.

Considerable lack of trueness is seen for the logistic and log–log models with bias values of close to −50, implying that the I_T and consequently I_A indices are 0.

This bias may be due to a matrix effect between the calibration and validation standards: the calibration standards are provided in buffer but the validation standards are in a complex matrix (homogenised eggs).

3.3. Precision

Two sets of precision values were calculated: repeatability and intermediate precision. Precision was better for the weighted $(1/X^2)$ quadratic regression model than for other quadratic models (Table 1). Considerable variability however was seen at the lowest concentration (7 ng/ml). The relative standard deviation (R.S.D.) of repeatability and the R.S.D. of the intermediate precision are shown in Table 2 for three models. These findings show acceptable repeatability but the intermediate precision which did not comply with the limits set in Decision 2002/657 [12].

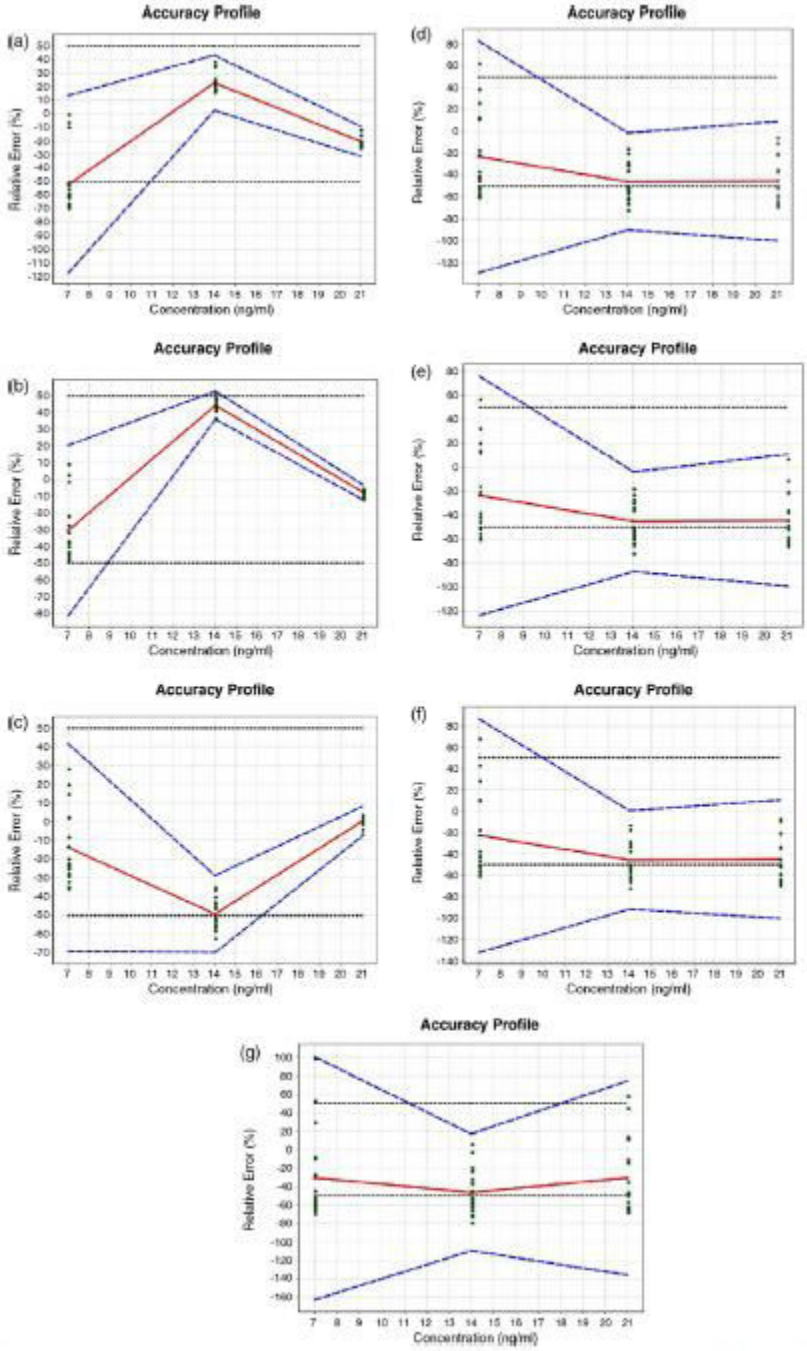


Fig. 1. Accuracy profile obtained for the measurement of the level of nicarbazin in eggs with (a) weighted ($1/X^2$) quadratic regression, (b) weighted ($1/X$) quadratic regression, (c) unweighted quadratic regression, (d) weighted four parameter logistic regression, (e) unweighted four parameter logistic regression, (f) weighted five parameter logistic regression, (g) log–log regression. Plain lines are the relative bias, the dashes lines are the β-expectations tolerance limits, the dotted curves represent the acceptance limits, the dots represent the relative back-calculated concentrations of the validation standards.

Table 1
Indices of the different regression models tested ranked by accuracy index (I_A).

Model	I_A	I_{DR}	I_P	I_T	Figure
Weighted ($1/X^2$) quadratic regression	0.64	0.72	0.70	0.52	1a
Weighted ($1/X$) quadratic regression	0.56	0.48	0.61	0.61	1b
Unweighted quadratic regression	0.50	0.34	0.57	0.65	1c
Weighted four parameter logistic regression	0	0	0	0	1d
Unweighted four parameter logistic regression	0	0	0	0	1e
Weighted five parameter logistic regression	0	0	0	0	1f
Unweighted log–log regression	0	0	0	0	1g
Unweighted five parameter logistic regression	NC	NC	NC	NC	
Weighted linear regression	NC	NC	NC	NC	
Linear regression	NC	NC	NC	NC	

I_A: accuracy index; I_{DR}: dosing range index; I_T: trueness index; I_P: precision index; NC: not calculated.

Table 2
Precision (repeatability and intermediate precision) obtained for models used, by concentration level tested.

Models	Concentration level (ng/g)					
	7		14		21	
	Repeat. (R.S.D. %)	IP (R.S.D. %)	Repeat. (R.S.D. %)	IP (R.S.D. %)	Repeat. (R.S.D. %)	IP (R.S.D. %)
Weighted ($1/X^2$) quadratic regression	2.72	23.62	1.36	7.49	1.01	4.11
Weighted ($1/X$) quadratic regression	4.23	18.69	2.26	3.53	1.34	1.90
Unweighted quadratic regression	6.30	20.75	3.52	7.98	2.09	3.27

Repeat: repeatability, IP: intermediate precision, R.S.D.: relative standard deviation (precision) (%).

3.4. Solutions tested

A number of solutions can be used to resolve the trueness problem. The first is to use a correction coefficient as described by Hubert et al. [19] to correct the matrix effect. The correction coefficient was computed from the slope of the linear equation linking theoretical spiked concentration to recovered concentration computed by inverse prediction. The correction coefficient used is the reciprocal of the slope achieved with the validation standards.

The equation of the line is:

$$[\text{Recovered}] = 1.867 + 0.454\ [\text{Added}]$$

and the correction coefficient to be applied to the instrument response is therefore: new results = (old results − 1.87)/0.45.

Fig. 2 shows the new accuracy profile then obtained. For trueness the I_T was close to 1, i.e. bias was close to 0 although precision was amplified at all levels. This solution cannot therefore be used.

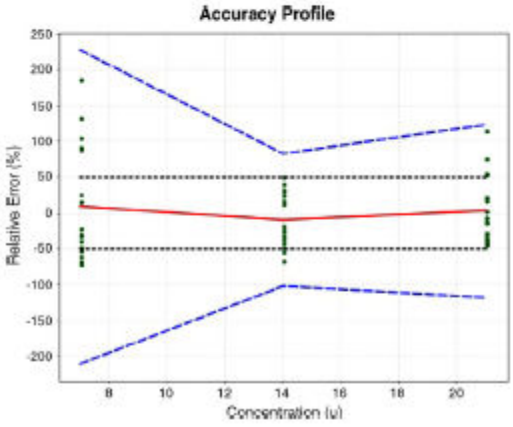


Fig. 2. Accuracy profile obtained with corrected results. Plain lines are the relative bias, the dashes lines are the β-expectations tolerance limits, the dotted curves represent the acceptance limits, the dots represent the relative back-calculated concentrations of the validation standards.

The second solution is to work in a dosing range from 14 to 21 ng/ml without correction. In this case the best response function is the weighted ($1/X$) quadratic regression.

The third solution is to use the method as a qualitative test.

3.5. Cut-off determination

A cut-off value was determined using blank samples and samples spiked to 70 μg/kg. Because the high variability seen in the validation and according to Decision 2002/657 [12] it is reasonable to set a false negative rate of 5%. A cut-off value of 20 μg/kg was established by calculating the 95th percentile of our samples. This value is acceptable as it is well below the DAL of 100 μg/kg. The false positive rate at this cut-off is less than 1%.

3.6. Comparison with initial data published

It is difficult to compare our validation results with those obtained by Huet et al. [8]. This author has not really calculated trueness and precision so the point by point comparison on these criteria is not possible. However, Huet has compared results obtained by an LC–MS–MS method and its ELISA method from incurred or spiked eggs and not from quantitative validation study with standard curves. The results have shown a high variability, i.e. accuracy, depending on concentration as shown by our validation results.

4. Conclusion

We have used the accuracy profile successfully to assess the performance of the ELISA kit for nicarbazin in eggs. The performance of the method was compared to the acceptance limits set in accordance with European Decision 2002/657 [12]. Under our conditions, however, the method cannot be validated for use as a semi-quantitative method as both trueness and precision are not within regulatory acceptable limits.

We have tested alternative solutions to solve the problem of trueness. The best solution found was to reduce the dosing range.

The method should be used as a qualitative method to detect the presence of nicarbazin in eggs with a cut-off of 20 μg/kg.

References

[1] L. Mortier, A.-C. Huet, C. Charlier, E. Daeseleire, P. Delahaut, C. van Peteghem, Food Addit. Contam. 22 (2005) 1120.
[2] Council Regulation (EC) No. 1831/2003, Off. J. Eur. Commun. L268 (2003) 29.
[3] Council Directive (EC) No. 70/524/EEC, Off. J. Eur. Commun. L270 (1970) 1.
[4] C.T. O'Keeffe, M. Capurro, E. Danaher, M. Campbell, K. Elliott, Food Addit. Contam. 24 (2007) 923.
[5] A. Cannavan, G. Ball, D.G. Kennedy, Food Addit. Contam. 17 (2000) 829.
[6] R. Young, A. Craig (Eds.), The Use and Misuse of Antibiotics in UK Agriculture—Part 3, Soil Association, Bristol, UK, 2001, p. 1.
[7] Council Regulation (EEC) No. 2377/90, Off. J. Eur. Commun. L224 (1990) 1.
[8] A.-C. Huet, L. Mortier, E. Daeseleire, T. Fodey, C. Elliott, P. Delahaut, Food Addit. Contam. Part A 22 (2005) 128.
[9] B. McCarney, L.M. Traynor, T.L. Fodey, S.R.H. Crooks, C.T. Elliott, Anal. Chim. Acta 483 (2003) 165.
[10] K. Campbell, T. Fodey, C. Flint Danks, M. Danaher, M. O'Keeffe, G. Kennedy, C.T. Elliott, J. Agric. Food Chem. 55 (2007) 2497.
[11] V. Hagren, S.R.H. Crooks, C.T. Elliott, T. Lovgren, M. Tuomola, J. Agric. Food Chem. 9 (2004) 2429.
[12] Commission Decision 2002/657/EC, Off. J. Eur. Commun. L221 (2002) 8.
[13] Ph. Hubert, P. Chiap, J. Crommen, B. Boulanger, E. Chapuzet, N. Mercier, S. Bervoas-Martin, P. Chevalier, D. Grandjean, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, Anal. Chim. Acta 391 (2) (1999) 135.
[14] Ph. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, STP Pharma 13 (2003) 101.
[15] Ph. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, J. Pharm. Biomed. Anal. 36 (3) (2004) 57.
[16] E. Chéneau, J. Henri, Y. Pirotais, J.P. Abjena, B. Roudaut, P. Sanders, M. Laurentie, J. Chromatogr. B 850 (2007) 15.
[17] B. DeSilva, W. Smith, R. Weiner, M. Kelley, J.M. Smolec, B. Lee, M. Khan, R. Tracey, H. Hill, A. Celniker, Pharm. Res. 20 (11) (2003) 1885.
[18] E. Rozet, V. Wascotte, N. Lecouturier, V. Préat, W. Dewé, B. Boulanger, Ph. Hubert, Anal. Chim. Acta 591 (2007) 239.
[19] Ph. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Laurentie, N. Mercier, G. Muzard, L. Valat, STP Pharma 16 (2) (2006) 87.

IV. ARTICLE 4

VALIDATION OF A FIVE PLATE TEST, THE STAR PROTOCOL, FOR THE SCREENING OF ANTIBIOTIC RESIDUES IN MUSCLE FROM DIFFERENT ANIMAL SPECIES ACCORDING TO THE EUROPEAN DECISION 2002/657/EC

Valérie GAUDIN, Céline HEDOU, Annie RAULT, Eric VERDON

Food Additives and Contaminants: Part A

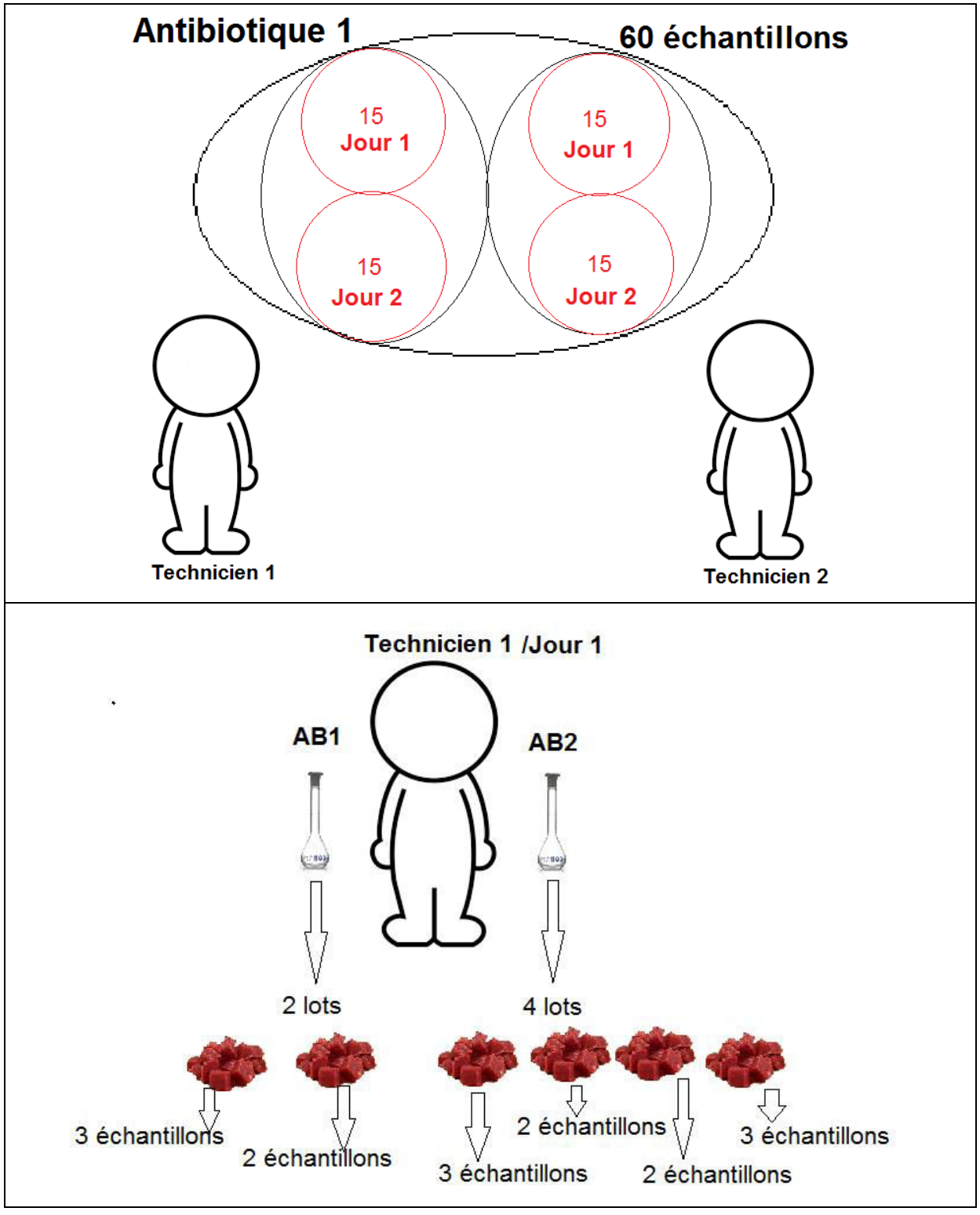
Contexte

Depuis des décennies, les méthodes microbiologiques sont couramment utilisées pour le dépistage des résidus d'antibiotiques dans les matrices d'origine animale, telles que le muscle, le rein, le lait, etc. La méthode nommée « Screening Test for Antibiotic Residues » (STAR) a été développée par notre laboratoire dans le cadre de nos activités de laboratoire de référence de l'union européenne (LRUE) pour les résidus d'antibiotiques. L'objectif de ce développement était l'amélioration des performances des méthodes microbiologiques et l'harmonisation des méthodes au niveau européen. Cette méthode dite « 5 boites » est basée sur la combinaison de 5 germes (*Bacillus cereus* (Bc6), *Bacillus stearothermophilus* (Bst), *Escherichia coli* (Ec8), *Micrococcus luteus* (Ml8) et *Bacillus subtilis* (Bs8)) inoculés dans 5 milieux différents par leur pH et leur composition. Cette combinaison permet de détecter plusieurs familles d'antibiotiques, grâce à l'inhibition de la croissance des bactéries en présence du résidu dans le muscle. La méthode STAR a été précédemment validée dans le lait [29], suivant la décision Européenne 2002/657/CE [13]. Nous avons alors décidé de valider cette méthode pour le muscle, suivant cette même décision. Toutefois, l'approche de la validation a été un peu différente et plus complexe, en raison de la spécificité du muscle qui est une matrice solide.

Méthodologie/Principaux résultats

Des échantillons de muscle de différentes espèces animales (porc, bovin, ovin, volaille) ont été préparés. Il s'agissait de muscle broyé grossièrement, qui ont été supplémentés avec une quantité connue d'antibiotique. Ensuite, le mélange a été homogénéisé avec un agitateur rotatif pendant 15 minutes, puis broyé plus finement. Enfin, les tissus « simulés » ont été aliquotés, puis conservés au congélateur à -20°C pendant un mois au maximum avant analyse. Dans un premier temps, 16 antibiotiques différents, appartenant à 9 familles d'antibiotiques (macrolides, aminosides, quinolones, tétracyclines, pénicillines, céphalosporines, sulfamides, lincosamides et divers), ont été testés à trois concentrations dans les tissus « simulés » (la LMR, et en fonction du niveau de détection attendu, 2 fois la LMR ou même plus). Cette phase préliminaire a permis de définir les concentrations cibles pour la validation. Les capacités de détection CCβ de 16 antibiotiques ont été déterminées dans le muscle de porc, en analysant 60 échantillons de muscle supplémentés (tissus « simulés »), provenant de lots différents. La préparation des tissus supplémentés a été réalisée par 2 techniciens, en suivant le protocole détaillé à la **Figure 48**. Ce protocole a été déterminé afin d'introduire le maximum de variabilité dans la préparation et l'analyse des échantillons (techniciens, espacement dans le temps (conditions ambiantes, lots de standards, de milieu gélosé, de bactéries, etc), lots différents de muscle de porc, solutions filles différentes), jours différents), pour obtenir des CCβ au plus près des conditions de routine. De plus, 49 lots différents de muscle de porc blancs (dépourvus de résidus d'antibiotiques) ont été analysés afin de déterminer la spécificité de la méthode. Toutes les analyses ont été réalisées en aveugle et tous les échantillons ont été déposés sur les cinq boites de la STAR. La méthode STAR avait été appliquée pendant 5 ans pour le dépistage des résidus d'antibiotiques dans le muscle de différentes espèces, en parallèle de la méthode des 4 boites (méthode officielle française).

Figure 48. Organisation de la préparation des échantillons pour la validation de la méthode STAR dans le muscle.

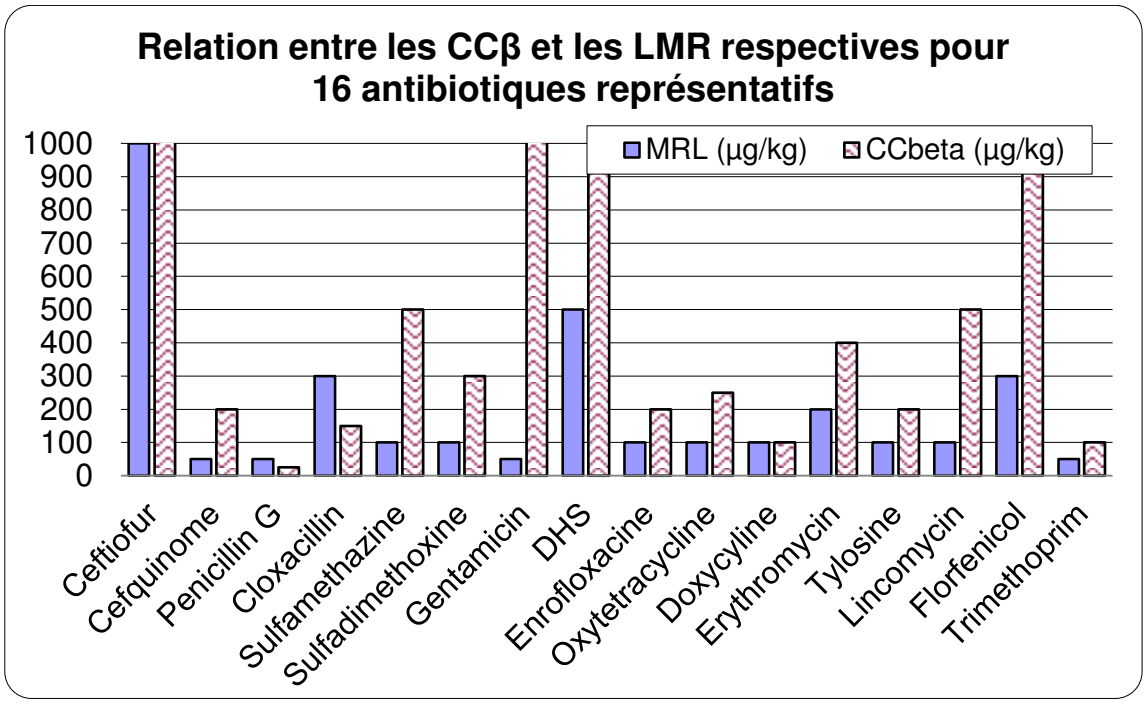


Toutefois, l'applicabilité de la méthode à d'autres espèces animales que le porc a été testée. Dans ce but, nous avons déterminé la spécificité et les CCβ de 5 antibiotiques (pénicilline G à 25 µg kg^{-1}, doxycycline à 100 µg kg^{-1}, érythromycine à 400 µg kg^{-1}, gentamicine à 6000 µg kg^{-1}, enrofloxacine à 200 µg $kg^{-1)}$, à partir de l'analyse de 20 échantillons de 3 espèces différentes, pendant 5 jours. Enfin, une étude de robustesse a été réalisée afin de tester les paramètres analytiques qui pourraient avoir une influence sur le résultat.

Le taux de faux-positifs de la méthode STAR était de 4 %, dans le muscle de porc, seulement sur la boite *Bst* (0 % sur les autres boites). Bien qu'il n'existe pas de critère maximal pour le taux de faux-positifs dans la décision 2002/657/CE, un taux inférieur à 5 % pour une méthode microbiologique de dépistage est généralement considéré comme satisfaisant. Les CCβ de 3 antibiotiques (pénicilline G, cloxacilline et doxycycline) sont inférieurs ou égaux à leurs LMR respective (**Figure 49**). Les CCβ de 8 antibiotiques (tylosine, sulfadiméthoxine, oxytétracycline, triméthoprime, érythromycine, ceftiofur, enrofloxacine et cefquinome) sont compris entre 1,5 et 3 fois leur LMR respective. Enfin, les CCβ des 5 derniers antibiotiques (lincomycine, gentamicine, dihydrostreptomycine, et florfénicol) sont supérieurs à 4 fois leur LMR respective. Chaque boite est spécifique de la détection d'une famille d'antibiotiques, excepté si les concentrations en antibiotiques sont élevées. L'applicabilité de la méthode STAR pour l'analyse de muscles d'autres espèces a été prouvée. En effet, la zone d'inhibition moyenne obtenue avec les autres espèces était comprise dans un intervalle de plus ou moins 25 % par rapport à celles obtenues avec le muscle de porc, ce qui était notre critère pour valider l'applicabilité. Un seuil de positivité plus élevé est appliqué à la boite Bst (4 mm au lieu de 2 mm d'inhibition pour les autres boites), car les échantillons blancs donnent des zones d'inhibition plus élevées sur cette boite que sur les autres. Deux résultats faux-positifs ont été obtenus sur la boite Bst, sur 30 échantillons analysés (bovin, volaille, ovin). Aucun résultat faux-positif n'a été observé sur les 4 autres boites, quelle que soit l'espèce animale.

Concernant la robustesse, quelle que soit la boite, aucun effet significatif de 3 facteurs testés (concentrations en bactéries, quantité de milieu, et temps d'incubation) n'a été démontré sur le taux de faux-négatifs et de faux-positifs, excepté pour la boite *Bst* avec la sulfadiméthoxine. L'augmentation de la concentration de bactéries et de la quantité de milieu pour cette boite augmente le taux de faux-négatifs, ce qui est bien connu pour ce type de méthodes. Enfin, seul le temps de pré-incubation a un impact sur plusieurs boites. Une pré-incubation d'une heure à température ambiante améliore les limites de détection de certaines boites (*Ec8*, *Bst* pour les sulfamides et *Kv8*).

Figure 49. Relation entre les CCβ et les LMR respectives pour 16 antibiotiques représentatifs.



Les 16 antibiotiques représentatifs sont représentés sur l'axe des x. Les CCβ pour les 16 antibiotiques représentatifs testés et leurs LMR respectives sont représentés sur l'axe des y. LMR = Limite Maximale de Résidu.

Conclusions/Importance

En conclusion, la méthode STAR est une méthode robuste, simple, peu couteuse, et qui permet le dépistage de 9 familles d'antibiotiques dans le muscle de différentes espèces animales. Toutefois, la méthode STAR n'est pas applicable pour le dépistage de la famille des aminosides et du florfénicol au niveau des LMR, de même pour certaines molécules de la famille des sulfamides, des macrolides, des quinolones et des tétracyclines.

La validation d'une méthode microbiologique de type boites est très lourde et longue pour plusieurs raisons :

- Les substances à LMR dans le spectre de ces méthodes sont très nombreuses (*eg.* plus de 50 dans le muscle).
- Un échantillon supplémenté ne peut pas contenir un mélange d'antibiotiques car ce type de méthode n'est pas spécifique. Donc le nombre d'échantillons à analyser est un multiple du nombre d'antibiotiques pour lequel on va déterminer un CCβ. De plus, si on mélangeait des antibiotiques, il y a un risque d'effet synergique ou antagoniste sur la croissance bactérienne.
- Du fait des résultats qualitatifs, le nombre d'échantillons nécessaires pour déterminer les capacités de détection CCβ avec une certitude statistique, est le plus souvent plus élevé que pour les méthodes de confirmation, excepté quand la concentration cible est au niveau ou en dessous de la LMR.
- Les matrices solides, comme le muscle, rendent la validation encore plus complexe. La production de muscles naturellement chargés nécessite des phases animales couteuses. De plus, obtenir une concentration cible dans la matrice (si elle est utilisée intacte pour la validation) s'avère complexe, en raison e la grande variabilité entre les animaux. De plus, l'homogénéité de muscles intacts est difficile à garantir puisque le principe actif ne se répartir pas de façon homogène dans le muscle. Enfin, réaliser une validation de méthodes sur la base de disques supplémentés en antibiotiques, en absence de la matrice, ne donne qu'une indication partielle des performances de la méthode dans les conditions réelles de la matrice. C'est pourquoi nous avons produit des tissus dits « simulés » pour réaliser la validation. L'utilisation de tissus « simulés » tente de s'approcher au plus près des performances réelles de la méthode, mais le muscle broyé n'est pas la matrice utilisée pour l'analyse d'échantillons en routine. Donc les tissus simulés peuvent produire des biais par rapport à des tissus entiers.

La décision européenne 2002/657/CE donne très peu d'informations pour la validation des méthodes de dépistage et surtout aucune recommandation technique pour la mise en place de la validation. C'est pourquoi le guide Européen pour la validation des méthodes de dépistage [256], qui aide à l'application de la décision 2002/657/CE [13] pour les méthodes de dépistage, a été rédigé par trois LRUE, en lien avec la DGSANTE. Ce guide recommande l'utilisation de tissus simulés et propose de restreindre la validation au minimum à une liste de substances représentatives. Cette liste n'est pas figée et le choix des antibiotiques se fait en fonction des antibiotiques détectés par la méthode, des substances ayant une LMR dans la matrice d'intérêt, des usages particuliers dans certains pays et des connaissances préalables sur les limites de détection de la méthode. Ces deux points ont constitué une grande avancée dans la validation de ce type de méthodes. Toutefois, valider ce type

de méthodes microbiologiques représente quand même beaucoup de temps et d'effort, quelle que soit la matrice d'intérêt. De ce fait, le guide Européen propose que cette validation dite initiale soit réalisée dans un seul laboratoire (le développeur ou un autre). Ensuite, les résultats de la validation doivent être mis à disposition d'un laboratoire dit récepteur qui veut mettre en place la méthode. Ce deuxième laboratoire pourra faire une validation réduite, dite de transfert (moins d'échantillons, pas d'étude de robustesse). Les caractéristiques de performance (*eg*. CCβ) obtenues dans le laboratoire récepteur seront alors comparées avec celles obtenues lors de la validation initiale. Ces performances devront être équivalentes. Enfin, la participation du laboratoire récepteur à des tests inter-laboratoires d'aptitude permettra de confirmer son aptitude à mettre en place la méthode dans son laboratoire.

Food Additives and Contaminants
Vol. 27, No. 7, July 2010, 935–952

Taylor & Francis
Taylor & Francis Group

Validation of a Five Plate Test, the STAR protocol, for the screening of antibiotic residues in muscle from different animal species according to European Decision 2002/657/EC

V. Gaudin*, C. Hedou, A. Rault and E. Verdon

Community Reference Laboratory, AFSSA Fougères – LERMVD, La Haute Marche – BP 90203, F-35302 Fougere Cedex, France

(*Received 20 August 2009; final version received 4 February 2010*)

The STAR protocol is a Five Plate Test (FPT) developed several years ago at the Community Reference Laboratory (CRL) for the screening of antimicrobial residues in milk and muscle. This paper presents the validation of this method according to European Decision 2002/657/EC and to an internal guideline for validation. A validation protocol based on 'simulated tissues' and on a list of 16 representative antimicrobials to be validated was implemented in our laboratory during several months for the STAR protocol. The performance characteristics of the method were determined (specificity, detection capabilities CCβ, applicability, ruggedness). In conclusion, the STAR protocol is applicable to the broad-spectrum detection of antibiotic residues in muscles of different animal species (pig, cattle, sheep, poultry). The method has good specificity (false-positive rate = 4%). The detection capabilities were determined for 16 antibiotics from different families in relation to their respective maximum residue limit (MRL): beta-lactams (penicillins and cephalosporins $\leq$ MRL), tetracyclines ($\leq$MRL and $\leq$2.5 MRL), macrolides (2 MRL), quinolones ($\leq$2 MRL), some sulphonamides ($\leq$3 MRL), and trimethoprim (2 MRL). However, the sensitivity of the STAR protocol towards aminoglycosides (>8 MRL) and florfenicol ($\leq$10 MRL) was unsatisfactory (>>MRL). The two objectives of this study were met: firstly, to validate the STAR protocol according to European Decision 2002/657/EC, then to demonstrate that the validation guideline developed to implement this decision is applicable to microbiological plate tests even for muscle. The use of simulated tissue appeared a good compromise between spiked discs with antibiotic solutions and incurred tissues. In addition, the choice of a list of representative antibiotics allowed the reduction of the scope of the validation, which was already costly in time and effort.

Keywords: in-house validation; screening – microbial screening; residues; antibiotics; antimicrobials; veterinary drugs; meat

Introduction

Due to the preventive or curative treatment of livestock, the presence of antibiotic residues could be found in food of animal origin. Traces of these antibiotics could cause various problems: problems of technological processing (e.g., milk), allergies to antibiotics following the ingestion of contaminated food (e.g., penicillins), antibiotic resistance of bacteria in humans, which could be transferred to pathogenic bacteria for humans. Therefore, maximum residue limits (MRLs) were set for the antibiotics approved for use in veterinary medicine for livestock. European Regulation No. 470/2009 of 6 May 2009 (European Commission 2009) establishes procedures for the establishment of MRLs of pharmacologically active substances in foodstuffs of animal origin and repeals Regulation No. 2377/90 of 26 June 1990 (European Commission 1990). To monitor the presence of residues, the first step is the screening step, which is to conclude whether or not a sample contains antibiotic residues at or above the MRL, in the case of permitted substances. Then, in case of positive screening, it is necessary to use physicochemical methods for the confirmation of identity and the quantification of the substance (European Commission 2002). The screening step is often based on microbiological screening methods that are cheap, easy to perform, and do not need specific and expensive equipment. These methods were used for many years. Some of them were developed even before the establishment of the MRLs (European Commission 1990) and many of them before the implementation of the rules for the validation of analytical methods (European Commission 2002).

The STAR protocol (for Screening Test for Antibiotic Residues), developed at the Community Reference Laboratory (CRL) for antimicrobial residues in food (AFSSA Fougeres, France), is intended for the qualitative detection of residues of substances with antimicrobial activity in milk and muscle by using

*Corresponding author. Email: v.gaudin@fougères.afssa.fr

ISSN 0265-203X print/ISSN 1464-5122 online

DOI: 10.1080/19440041003677483
http://www.informaworld.com

bacterial strains sensitive to antibiotics. This method is based on five different plates (Five Plate Test), dedicated to the detection of specific families of antibiotics. The first validation of the STAR protocol was organized in 1999 by way of a collaborative study with spiked discs, blank muscles and incurred muscles (Fuselier et al. 2000). Seven antibiotics from six families were tested. The detection levels were included between 1 and 4 MRLs, but it was dependent on the production of incurred materials. These results were promising for the implementation of the STAR protocol. The STAR protocol was then validated for the screening of antibiotics in milk (Gaudin et al. 2004). Sensitivity was established by the analysis of milk samples spiked with 66 antibiotics at eight different concentrations. Ten different groups of antibiotics were studied: macrolides, aminoglycosides, cephalosporins, penicillins, quinolones, tetracyclines, sulphonamides, lincosamides, phenicols and miscellaneous drugs. The STAR protocol was able to detect 21 antibiotics at or below the MRL, and a further 27 drugs could be detected at levels from the MRL up to four times the MRL. The STAR protocol was at least twice as sensitive as conventional methods for macrolides, quinolones and tetracyclines. Each plate was preferentially sensitive for one or two families of antibacterials: the plate *Bacillus cereus* for tetracyclines, the plate *Escherichia coli* for quinolones, the plate *Bacillus subtilis* for aminoglycosides, the plate *Kocuria rhizophila* for macrolides and the plate *Bacillus stearothermophilus* for sulphonamides and beta-lactams.

Since 2002, every analytical method used for the analysis of monitoring routine samples has to be validated according to European Decision 2002/657/EC (European Commission 2002), which concerns the performance of analytical methods and the interpretation of results. The level of validation of confirmatory methods is now quite satisfactory. However, very little information is contained in that decision concerning the validation of screening methods. Two main information are contained in the decision: firstly, which performance characteristics have to be determined for a screening method and; secondly, the following information about the detection capability required for screening methods. Screening methods are:

> only those analytical techniques, for which it can be demonstrated in a documented traceable manner that they are validated and have a false compliant rate of <5% (β-error) at the level of interest shall be used for screening purposes in conformity with Directive 96/23/EC (European Commission 1996). In the case of a suspected non-compliant result, this result shall be confirmed by a confirmatory method.

Finally, the difficulty is how to use this information to implement a validation for a screening method.

Moreover, the validation of microbiological screening methods presents specific issues, different from physicochemical methods which are often specific methods. Firstly, the number of antibiotics to be validated is very high because all the antibiotics having an MRL in the corresponding matrix should be validated. It could be also the case for newly developed multi-residue methods by liquid chromatography-tandem mass spectrometry (LC-MS/MS). However, microbiological methods do not allow one to identify the antibiotic residue present in the sample. Therefore, each antibiotic has to be tested independently and the number of analyses increased. Secondly, these kind of methods are only qualitative methods giving a response as negative, positive or doubtful. Deciding on a sample size for a qualitative enquiry can be even more difficult than for a quantitative enquiry because there are no definite rules to be followed. In general, sample size depends on the nature of the analysis to be performed and on the desired precision of the estimates one wishes to achieve. The larger the sample size, the more sure one can be that the answers truly reflect the population. This indicates that for a given confidence level (e.g., 95%; β error = 5%), the larger the sample size (n), the smaller the confidence interval (interval estimate of a population parameter). For $n = 20$ and a percentage of answers of 50% (i.e., 50% answers positive; 50% answers negative), the confidence interval is 22%. For $n = 60$ and a percentage of answers of 50% (50% positive; 50% negative), the confidence interval is 13%. A greater sample is then required to decrease the confidence interval and therefore the chance of error. Therefore, from a statistical point of view, the number of samples to be analysed should be higher than for quantitative methods. Thirdly, spiking liquid matrices (e.g., milk, juice meat) are easy and validation could be performed on spiked samples. The problem is specific to solid matrices like muscle. In fact, plate tests are based on the analysis of raw muscle (slices of meat). Therefore, the validation should be conducted on blank and incurred raw muscle. However, the production of incurred materials for each antimicrobial at one concentration would be time and money consuming. Consequently, very few microbiological methods were validated in the matrix, especially in muscle and other solid matrices according to Decision 2002/657/EC (European Commission 2002). A new guideline document supplements Decision 2002/657/EC regarding the validation of screening methods. The CRL in Fougères, in collaboration with the CRL in Berlin and in agreement with the CRL in Bilthoven, and after consultation through the NRL (National Reference Laboratories) network, has drafted this document with the purpose of assisting residue laboratories to validate screening methods. This document is now finalized and officially published on the DGSANCO website since

21 January 2010 (Anon 2010). This guideline deals with the initial validation and also a shortened or 'abridged' validation, which under certain conditions allows for the transfer of methods already validated in one laboratory to a second one. The guideline proposes some recommendations to implement a validation protocol for screening methods. It explains the performance characteristics to be determined (specificity, detection capability, robustness, etc.) and how to determine them in practice. Moreover, the number of samples necessary to validate a screening method is discussed. Finally, some new concepts were introduced in the guideline: the preparation of 'simulated tissues' and a list of representative substances to be validated. These two concepts will be detailed below in the validation protocol.

The validation conducted in this study is based on this validation guideline which is also the internal guideline in our laboratory for the validation of screening methods. Two objectives were set: firstly, to validate the STAR protocol for its application to the detection of antibiotic residues in muscles from different animal species according to Decision 2002/657/EC (European Commission 2002); and secondly, to show that the European guideline for the validation (Anon. 2010) and based on Decision 2002/657/EC (European Commission 2002) was applicable to the validation of a microbiological screening method.

This paper will present the validation protocol and then the results of the validation of the STAR protocol for the screening of antibiotic residues in muscles of different animal species.

Material and methods

Chemicals and standard solutions

Antibiotic and sulphonamide standards were provided by Sigma (Lyon, France), except cefquinome (Intervet, Angers, France), ceftiofur (Upjohn, Val-de-Reuil, France), enrofloxacin (Bayer, Puteaux, France), and doxycycline (Virbac, Carros, France).

STAR protocol

The STAR protocol is a Five Plate Test already published for the analysis of milk (Gaudin et al. 2004). Five test organisms – *Bacillus subtilis* B.G.A spores (Merck), *Kocuria rhizophila* ATCC 9341, *Bacillus cereus* ATCC 11778, *Escherichia coli* ATCC 11303 (all Pasteur Institute, France), and *Bacillus stearothermophilus* ATCC 10149 (Merck) – were inoculated in five different media. The five following culture media were used respectively: Antibiotic medium II at pH 8.0 (plate Bs8), test agar at pH 8 (Merck) (plate Kv8), test agar at pH 6 (Merck) (plate Bc6), test agar at pH 8 (Merck) (plate Ec8), and Diagnostic Sensitive Test (DST) (Oxoid commercialized by Unipath Ltd, Basingstoke, UK) (plate Bst). Culture media were prepared as recommended by the supplier and sterilized. Then, 5 ml of inoculated medium were added to a Petri dish placed on a cold horizontal surface. In routine use, a cylindrical plug of 8 mm in diameter and 2 cm long is cut in frozen muscle using a cork borer. Slices of muscle samples of 2 mm in thickness are then cut and placed on the plates. The same protocol was applied to 'simulated tissues'. Finally, the plates are incubated: at 30°C for at least 18 h for Bs8 and Bc6, at 37°C for at least 24 h for Kv8, at 37°C for at least 18 h for Bc6, and at 55°C for 15 to 16 h for Bst.

A muscle sample was considered positive when the inhibition zone (IZ) around the meat sample was equal or superior to 2 mm in width on plates Bs8, Kv8, Bc6 and Ec8 and/or the IZ equal or superior to 4 mm in width on plate Bst.

Since 2004, the medium test agar pH 7.2 was replaced by antibiotic medium II at pH 8.0 (Difco, reference no. 259310) because it was not commercialized anymore. Moreover, the positive control of plate Bst was modified for practical reasons: sulfamethazine at 1000 $\mu g l^{-1}$ was replaced by amoxicillin at 40 $\mu g l^{-1}$. Positive controls consist of 30 µl of antibiotic solutions which are put on paper discs of 9 mm diameter (Durieux, France). There is a specific positive control for each plate: streptomycin at 2000 $\mu g l^{-1}$ on plate Bs8, tylosin at 1000 $\mu g l^{-1}$ on plate Kv8, oxytetracycline at 800 $\mu g l^{-1}$ on plate Bc6, enrofloxacin 800 $\mu g l^{-1}$ on plate Ec8, and amoxicillin at 40 $\mu g l^{-1}$ on plate Bst. The validity of each day of analysis depends on the results of the positive controls that have to be included in the following intervals: 5.5 ± 1.5 mm, 6.5 ± 1.5 mm, 6.0 ± 1.5 mm, 7.0 ± 1.5 mm and 6.0 ± 1.5 mm for Bs8, Kv8, Bc6, Ec8 and Bst respectively.

Validation protocol

Simulated tissues

In 2002, we studied the sensitivity of 35 antibiotics by way of antibiotic-spiked discs. However, this way of working was not completely satisfactory because the interference of the muscle matrix was absent. In fact, Okerman, De Wasch et al. (1998) showed that the tissue matrix has an effect on the sensitivity of the test plates. During that study, pieces of frozen meat laid on paper discs impregnated with antibiotic standard solutions were used for validation. Usually, IZ decreased when spiked meat samples were analysed, compared with antibiotic spiked discs without meat. The same conclusions were reported by Pikkemaat et al. (2007). Because of the difficulty and the high cost of production of the incurred materials for validation of a microbiological plate test, it was decided to work on what we called 'simulated tissue'. Furthermore,

it was impossible to mix several antibiotics in the same sample, since the method was not specific and did not identify the molecule that produced an inhibitory effect. Moreover, one could observe a cumulative effect of antibiotics in their inhibitory activity when several antibiotics are contained in one sample.

Therefore, a preliminary study was conducted to determine what was the best way to prepare simulated tissues and what kind of preparation would give the closest result to the actual samples. This study was based on the experience of two National Reference Laboratories (Vicente Calderon, AESAN, Spain; Anna Liiza Myllyniemi, EVIRA, Finland) that had already worked on such samples. During this preliminary study, the way of preparation of simulated tissue was also tested for homogeneity of the material. Regarding the results, the homogeneity was satisfactory.

Production of spiked materials ('Simulated tissue')

A preliminary study was conducted to determine the concentrations of antibiotics to be spiked in muscle samples for later validation. Sixteen different antibiotics were tested at various concentrations (the MRL and according to the assumed detection limit at half MRL, two times the MRL or more) in simulated tissues. The tested concentrations were based on the results of the validation of the STAR protocol in milk (Gaudin et al. 2004). These tests were repeated several times. The concentration that always gave positive results (IZ higher than 2 mm, generally between 3 and 4 mm) and if possible concentrations lower than or equal to the respective MRLs were chosen.

The antibiotics and the corresponding concentrations chosen for the validation are presented in Table 1. Sixteen different antibiotics were used in this study. Nine different groups of antibiotics were studied: macrolides, aminoglycosides, cephalosporins, penicillins, quinolones, tetracyclines, sulphonamides, lincosamides and miscellaneous drugs. Stock solutions of the 16 different antimicrobials were prepared at a concentration of 1 mg ml^{-1}, after correction for potency. Working solutions were then prepared by dilutions in distilled water. Different batches of muscle were purchased in supermarkets. Muscle was first coarsely minced. To prevent antibiotic contamination, finely minced blank muscle samples were prepared first, on the same day, with the four different batches of muscle, to be tested in parallel with the spiked samples. Then 1 ml of working solution was added to 100 g of blank minced muscle. Spiked muscle was homogenized during 15 min in the same rotary hatcher. At the end, the meat was finely minced. Each muscle material was finally put in plastic bottles or plastic bags. Each material was codified by the director of the study and then frozen at −20°C. The 'simulated tissues' were always analysed in a maximum period of 1 month.

Specificity

A total of 49 batches of pork muscle of different origins were analysed in the end. Most of them (40 batches) were tested in blind duplicate on two different days and by two different technicians.

Detection capabilities CCβ

According to Decision 2002/657/EC (European Commission 2002), the detection capability (CCβ) of a method is defined as:

> the smallest content of the substance that may be detected, identified and/or quantified in a sample with an error probability of β. In the case of substances with an established permitted limit, this means that the detection capability is the concentration at which the method is able to detect permitted limit concentrations with a statistical certainty of $1-\beta$.

Given the expected detection capabilities for many antibiotics, which are often close to the MRLs, it was chosen to determine the detection capabilities by analysing 60 minced muscle samples spiked at one

Table 1. Chosen antibiotics and corresponding concentrations for the validation study.

Antibiotic family	Representative antibiotic	MRL (µg kg^{-1})	Chosen concentration (µg kg^{-1})
Cephalosporins	*Ceftiofur*/Cefquinome	1000/50	1500/200
Penicillins	Penicillin G/Cloxacillin	50/300	25/150
Sulfonamides	*Sulfamethazine*/*Sulfadimethoxine*	100/100	500/300
Aminoglycosides	Gentamicin (GTM)/*Dihydrostreptomycin (DHS)*	50/500	2000/4000
Quinolones	Enrofloxacin	100	200
Tetracyclines	*Oxytetracycline*/*Doxycycline*	100/100	250/100
Macrolides	Erythromycin/*Tylosin*	200/100	400/200
Lincosamides	Lincomycin	100	500
Phenicols	*Florfenicol*	300	3000
Miscellaneous	Trimethoprim	50	100

Note: Antibiotics shown in italics are those that were substitutes for the antibiotics initially included in the list of representative antibiotics.

antibiotic concentration, instead of 20 samples to reach the most statically significant determination of the CCβ (European Commission 2002).

The preparation of the samples is detailed in Figure 1. To introduce a maximum of variability in the determination of the detection capabilities, 30 samples were prepared and analysed by one technician and another 30 samples by other technician. In addition, the 30 samples were divided into two sets of preparation. In each set, 15 samples for the same antibiotic concentration were prepared and analysed by each technician. For each antibiotic and each technician, the two sets were spaced out to introduce variability in the period of analysis (ambient conditions) and therefore variability of batches (media, bacteria, antibiotic standard, etc.). Moreover, in each set of preparation of four antibiotics, four different batches of porcine muscle were used. Finally, on each day of preparation, two different working solutions were prepared from the stock solution of antibiotic and used to spike the different batches of muscle. The coding of the samples was performed by the head of the study to ensure that the analyses would be performed blindly, and then frozen at −20°C.

In the determination of specificity and detection capabilities, all samples were analysed on the five plates of the STAR protocol to check the specificity of the plates for the different families of antibiotics. Five days of analyses were performed with the STAR protocol for each set of preparations and each technician. Each day, 14 samples were analysed blindly (twelve spiked samples and two blank samples). Therefore, all blank and spiked samples were analysed twice or thrice, on 2 or 3 different days.

It must be underlined that the analysis of the 60 samples for each antimicrobial was performed step by step. When one false-negative result or more was obtained after the analysis of ten samples, either the validation was stopped at this concentration and started again with an increased concentration if it was of interest (e.g., near the MRL), or the validation went on with the same antibiotic and the same concentration because the concentration was already much higher than the concerned MRL.

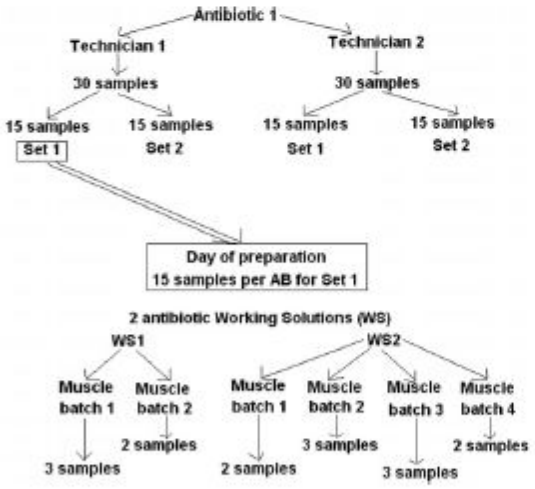


Figure 1. Preparation of the simulated tissues. Two technicians each prepared 30 samples per antibiotic (AB), divided in two sets of 15 samples each (Sets 1 and 2). The two sets were spaced out to introduce variability in the period of analysis (ambient conditions) and, therefore, variability of batches (media, bacteria, antibiotic standard, etc.). The example of preparation of Set 1 by technician 1 is presented here. Four different batches of porcine muscle were used. Two different working solutions (WS) were prepared from the stock solution of antibiotic and used to spike the different batches of muscle. Finally, 15 samples were prepared per antibiotic.

Determination of CCβ

After the analysis of the 60 spiked samples, the concentration level, where only less than 5% of false compliant results remain, was the detection capability CCβ of the method (three false compliant results in a maximum of 60 spiked samples).

Applicability study

The STAR protocol has been implemented in parallel with the Four Plate Test for the analysis of muscles from different animal species during the routine analysis of field samples (cattle, sheep, poultry, etc.) in National Monitoring Plans for at least 5 years. The positive samples at the screening step are then confirmed by a multi-residue LC-MS/MS method (data to be published). Therefore, we have a lot of experience concerning the applicability of the method to muscles from different animal species.

However, to complete this study, we decided to perform a study to determine the CCβ of some antibiotics in cattle, sheep and poultry muscles, antibiotics for which the CCβ were determined previously in porcine muscle. Therefore, the CCβ of five antimicrobials (penicillin G at 25 $\mu g\,kg^{-1}$, doxycycline at 100 $\mu g\,kg^{-1}$, erythromycin at 400 $\mu g\,kg^{-1}$, gentamicin at 6000 $\mu g\,kg^{-1}$, and enrofloxacin at 200 $\mu g\,kg^{-1}$) were determined with 20 samples from each of the three species over 5 days and compared with the CCβ calculated for porcine muscle. In the applicability study, only the specific plate (which has presented IZ for pig muscle) was tested with the corresponding antibiotic.

The applicability would be proved if the CCβ determined for these species are similar to the CCβ determined for porcine muscle (the average IZ obtained for each species on the specific plate should

be similar: accepted deviation ±25%). A variability of 25% is usually accepted with microbiological methods (for positive controls or spiked muscles). Moreover, this range of variability was observed during the validation in porcine muscle. The specificity of the STAR protocol with other species was also studied by the analysis of different blank muscle samples.

Ruggedness study

The ruggedness of the method was partly evaluated during the validation study, since different batches of media, bacterial strains, antimicrobial standards were used. In addition, two technicians were involved in the study, analysing each half of the samples and the variability between technicians could be estimated.

In Decision 2002/657/EC (European Commission 2002), ruggedness is defined as 'the susceptibility of an analytical method to changes in experimental conditions [...] under which the method can be applied as presented or with specified minor modifications'. A ruggedness study based on an experimental design was carried out and thus allowed the testing of analytical parameters which were different from those tested during the first step of validation. A full factorial design was chosen for the analysis of the effects and interactions of four independent factors (Renard et al. 1992). The design matrix is presented in Table 2.

Factors that may influence the measurement results were selected: bacteria concentration (A), medium quantity (B) in the plate, incubation time (C) and pre-incubation time at room temperature (D). These factors were modified in an order of magnitude corresponding to the usual differences: factor A concentration in bacteria: ±30%; factor B quantity of medium 5 ± 0.5 ml; factor C incubation time ±10%; and factor D pre-incubation time (the period at room temperature before incubation in the incubator: 1 h of pre-incubation or no pre-incubation).

The ruggedness of the study was focused on six different representative antimicrobials which were specifically detected on each of the five plates on muscle: penicillin G at 25 µg kg^{-1} and sulfadimethoxine at 300 µg kg^{-1} on plate Bst, doxycycline at 100 µg kg^{-1} on plate Bc6, erythromycin at 400 µg kg^{-1} on plate Kv8, gentamicin at 6000 µg kg^{-1} on plate Bs8, and enrofloxacin at 200 µg kg^{-1} on plate Ec8. The chosen concentrations were equal to the detection capabilities determined in the first part of the validation study for each of these antibiotics, except for gentamicin. The samples were prepared from pig muscle only. All blank and spiked minced muscle samples were prepared the day before starting the ruggedness study, and were frozen at −20°C. The study was performed blindly (codified samples). Then on each day of analysis (each run), four different blank materials and four different spiked materials per antibiotic were analysed.

Results and discussion

Specificity

Specificity and detection capabilities were determined for pig matrix because it is the most common species analysed in the laboratory. A total of 49 different batches of porcine muscle were tested. Few variations were observed. Of 176 analyses in total, only seven gave false-positive results (4%). These seven false-

Table 2. Design matrix of the ruggedness study.

Run	Levels	A[a]	B[b]	C[c]	D = ABC[d]	AB + CD	AC + BD	BC + AD
1	+	−[e]	−	−	−	+	+	+
2	+	+	−	−	+	−	−	+
3	+	−	+	−	+	−	+	−
4	+	+	+	−	−	+	−	−
5	+	−	−	+	+	+	−	−
6	+	+	−	+	−	−	+	−
7	+	−	+	+	−	−	−	+
8	+	+	+	+	+	+	+	+

Notes: Run 1 = day 1.
[a]Concentration of bacteria.
[b]Medium quantity.
[c]Incubation time.
[d]Pre-incubation time.
[e]Each factor was modified in an order of magnitude corresponding to the usual differences: factor A concentration in bacteria, ±30%; factor B quantity of medium, 5 ± 0.5 ml; factor C incubation time, ±10%; factor D pre-incubation time (period at room temperature before incubation in the incubator), 1 h of pre-incubation or no pre-incubation.
+, Increasing of the factor (i.e., +30%); −, decreasing of the factor (i.e., −30%). AB + CD, AC + BD and BC + AD are the evaluation of the impact of the combination of the different factors.

positive results were obtained with five different muscle batches (two false-positive results for two batches and one false-positive result for each of three batches). All false-positive results appeared on the plate Bst, none on the other plates. The specificity of the STAR protocol for the detection of antimicrobial residues in porcine muscle was very satisfactory.

Detection capabilities

The results presented in Table 3 are those obtained from 60 samples for each antibiotic at at least one tested concentration.

For some antibiotics, after 15 or 30 analyses by one or two technicians, either it turned out that too many false-negative results were obtained at the selected concentration or oppositely the first tested concentration gave higher IZ than anticipated, suggesting that the validation could be performed at a lower concentration. At these concentrations, fewer than 60 samples were analysed; however, it was helpful for the determination of the detection capability.

Figure 2 shows the relationship between detection capabilities calculated for the 16 representative antibiotics and their respective MRLs. It allows one to quickly visualize the difference between CCβ and MRLs. The respective MRLs ($\mu g\,kg^{-1}$) are represented in plain bars. The hatched bars represent the respective CCβ ($\mu g\,kg^{-1}$). Therefore, the greater the hatched area, the less sensitive is the STAR protocol for the corresponding antibiotic. On the contrary, when the plain zone and hatched zone are very close, the sensitivity of the method for the corresponding antibiotic is satisfactory. When the CCβ is higher than $1000\,\mu g\,kg^{-1}$, the value of the CCβ is written on the top of the corresponding bar.

As a conclusion, the detection capabilities of penicillin G, cloxacillin and doxycycline were equal to or lower than their respective MRLs. The detection capabilities of tylosin, sulfadimethoxine, oxytetracycline, trimethoprim, erythromycin, ceftiofur, enrofloxacin and cefquinome were between 1.5 and three times their respective MRLs. The detection capability of lincomycin was between four and five times its MRL. Finally, the detection capabilities of gentamicin (GTM) (>40*MRL), dihydrostreptomycin (DHS) (>8*MRL) and florfenicol (>10*MRL), were much higher than their respective MRLs. The STAR protocol is not suitable for the detection of the two aminoglycosides and probably not for the detection of the entire family, given the activity profiles obtained during the validation in milk (Gaudin et al. 2004). In fact, these activity profiles showed that GTM and DHS were the two aminoglycosides that were detected at the lowest concentrations, especially GTM. The sensitivities for GTM in milk ($MRL = 100\,\mu g\,l^{-1}$) and to DHS ($MRL = 200\,\mu g\,l^{-1}$) were equal to three and five times the respective MRLs. The sensitivities determined in milk for the other aminoglycosides were: neomycin $1500\,\mu g\,l^{-1}$ ($MRL = 1500\,\mu g\,l^{-1}$), streptomycin $1000\,\mu g\,l^{-1}$ ($MRL = 200\,\mu g\,l^{-1}$), kanamycin $1000\,\mu g\,l^{-1}$ ($MRL = 150\,\mu g\,l^{-1}$), spectinomycin $40{,}000\,\mu g\,l^{-1}$ ($MRL = 300\,\mu g\,l^{-1}$), paromomycin $2000\,\mu g\,l^{-1}$ (not authorized in milk), and apramycin $4000\,\mu g\,l^{-1}$ (not authorized in milk). Therefore, the sensitivities for GTM and DHS were much better in milk than in muscle.

During the determination of detection capabilities in porcine muscle, it was demonstrated that the STAR protocol could help in confirming the identity of a family of molecules present in the sample. Table 4 shows the plates on which each antibiotic reacted preferentially. Some antibiotics (e.g., cefquinome, ceftiofur, doxycycline, enrofloxacin and florfenicol) caused inhibitions on several plates at the tested concentrations (including CCβ).

The results are obviously similar to the global orientation given in the STAR protocol: beta-lactams and sulphonamides were detected preferentially on plate Bst, tetracyclines on plate Bc6, quinolones on plate Ec8 and macrolides on plate Kv8. However, there were two exceptions. The aminoglycosides were detected on plate Bst with the concentrations tested for the determination of CCβ, whereas for higher concentrations than CCβ they were detected on plate Bs8 (=Bs7.2). However, when the validation of the STAR protocol was performed in milk, Bs8 was the specific plate for the detection of aminoglycosides. Secondly, tylosin was detected preferentially on Bst instead of Kv8 at this concentration, which had already been observed during the validation of the STAR protocol in milk. Therefore, if a positive result is reported on plate Bst, the confirmation should be directed onto beta-lactams, sulphonamides, but also tylosin and aminoglycosides.

Applicability study

Given the unsatisfactory results for aminoglycosides (>>MRL), an additional study was conducted on GTM and DHS to choose one of these antibiotics and the good concentration for applicability and ruggedness studies. This study determined that the detection capability of GTM was around $4000\,\mu g\,kg^{-1}$ and DHS around $6000\,\mu g\,kg^{-1}$. Therefore, GTM at $6000\,\mu g\,kg^{-1}$ was chosen, a concentration higher than the estimated detection capability ($4000\,\mu g\,kg^{-1}$) due to reading difficulties (partial inhibition zone (PIZ) and regrowth in the IZ).

Figure 3 shows the average IZ on the five different plates (Bst, Bc6, Kv8, Ec8 and Bs8) for blank muscles of different animal species and muscles spiked with five different antibiotics (each reacted specifically

Table 3. Detection capabilities CCβ ($\mu g\ kg^{-1}$) for the 16 antimicrobials.

Antibiotic family	Antibiotic	MRL muscle ($\mu g\ kg^{-1}$)	Mean IZ[a] (mm) ± SD	Qualitative results	CCβ ($\mu g\ kg^{-1}$)
Penicillins	Penicillin G	50	9.7 ± 1.1	60+	≤25
	Cloxacillin	300	6.4 ± 0.2	56+/4− (at 150 $\mu g\ kg^{-1}$)	Between 150 and 300
Cephalosporins	Cefquinome	50	6.0 ± 0.3	14+/1D (at 100 $\mu g\ kg^{-1}$) 60+ (at 200 $\mu g\ kg^{-1}$)	Between 100 and 200
	Ceftiofur	1000	5.3 ± 0.7	56+/2D/2−	1500
Tetracyclines	Oxytetracycline (OTC)	100	5.1 ± 0.5	60+	≤250
	Doxycycline	100	5.4 ± 0.5	60+	≤100
Macrolides	Erythromycin	200	5.0 ± 1.6	56+/1D/3−	400
	Tylosin	100	5.1 ± 0.7	55+/3D/2−	200
Quinolones	Enrofloxacin	100	5.9 ± 1.9	8+/4D/3− (at 100 $\mu g\ kg^{-1}$) 60+ (at 200 $\mu g\ kg^{-1}$)	Between 100 and 200
Sulfonamides	Sulfadimethoxine (SDMX)	100	8.8 ± 1.3	12+/1D/2− (at 200 $\mu g\ kg^{-1}$) 59+/1D (at 300 $\mu g\ kg^{-1}$)	Between 200 and 300
	Sulfamethazine (SMZ)	100	5.4 ± 2.5	37+/2D/21−	>500
Aminoglycosides	Dihydrostreptomycin (DHS)	500	2.0 ± 0.7	24+/3D/33−	>4000
	Gentamicin (GTM)	50	4.3 ± 0.4	45+/3D/12−	>2000
Lincosamides	Lincomycin	100	5.8 ± 0.6	60+ (at 500 $\mu g\ kg^{-1}$)	Between 350 and 500
Miscellaneous	Trimethoprim (TMP)	50	7.9 ± 0.8	5+/10− (at 50 $\mu g\ kg^{-1}$) 59+ out of 59 (at 100 $\mu g\ kg^{-1}$)	Between 50 and 100
	Florfenicol	300	4.8 ± 0.2	8+/3D/19− (at 1000 $\mu g\ kg^{-1}$) 60+ (at 3000 $\mu g\ kg^{-1}$)	Between 1000 and 3000

Notes: [a]IZ, inhibition zone (mm); +, positive result; −, negative result; D, doubtful result (0 < IZ < 2 mm).
SD, standard deviation.

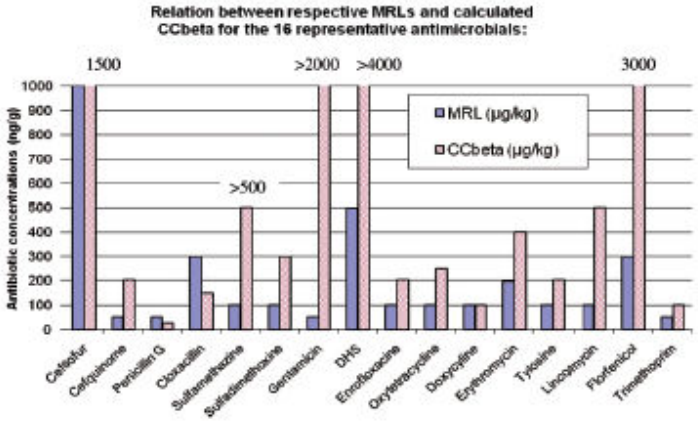


Figure 2. Relationship between the calculated detection capabilities and the respective MRLs of the 16 representative antibiotics. The 16 representative antibiotics are represented on the *x*-axis. The calculated CCβ of the 16 tested representative antimicrobials and the respective MRLs are represented on the *y*-axis. MRL, maximum residue limit.

Table 4. Specificity of the plates.

Antibiotic family	Antibiotic	Specific plate
Penicillins	Penicillin G	Bst
	Cloxacilline	Bst
Cephalosporines	Cefquinome	Kv8 (Bst and Ec8)
	Ceftiofur	Bst (and Ec8)
Tetracyclines	Oxytetracycline (OTC)	Bc6
	Doxycycline	Bc6 (and Bs8)
Macrolides	Erythromycin	Kv8
	Tylosin	Bst
Quinolones	Enrofloxacin	Ec8 (and Bst)
Sulfonamides	Sulfadimethoxine (SDMX)	Bst
	Sulfamethazine (SMZ)	Bst
Aminoglycosides	Dihydrostreptomycin (DHS)	Bst
	Gentamicin (GTM)	Bst
Lincosamides	Lincomycin	Bst
Miscellaneous	Trimethoprim (TMP)	Bst
	Florfenicol	Bs8 (Bc6 and Bst)

on one of the plates): penicillin G at 25 µg kg^{-1} on Bst (Figure 3a), doxycycline at 100 µg kg^{-1} on Bc6 (Figure 3b), erythromycin at 400 µg kg^{-1} on Kv8 (Figure 3c), enrofloxacin of 200 µg kg^{-1} on Ec8 (Figure 3d) and gentamicin at 6000 µg kg^{-1} on Bs8 (Figure 3e). The results for the porcine muscle were those obtained during the determination of CCβ for four plates (Bst, Bc6, Kv8, Ec8) and during the additional study on aminoglycosides for plate Bs8.

The overall mean inhibition zone inter-species was equal to 9.0 ± 1.3 mm, 5.2 ± 1.0 mm, 5.2 ± 1.4 mm, 6.1 ± 1.1 mm and 5.4 ± 0.9 mm for Bst, Bc6, Kv8, Ec8 and Bs8 respectively. Whichever was the tested antibiotic and the plate, the discrimination between blank and spiked samples was very clear and easy for each species. Moreover, the average IZ for cattle, sheep and poultry muscle was in the fixed interval of ±25% compared with the average IZ for porcine muscle for plates Bst, Bc6, Ec8 and Bs8. Concerning plate Kv8, the average IZ for cattle and sheep muscle was in the fixed interval of ±25% compared with the average IZ for porcine muscle. The poultry muscle was out of the interval for plate Kv8, but with an average IZ of 33% higher than the porcine muscle, so the sensitivity was better in poultry muscle.

Two blank samples of cattle, two sheep and two poultry were analysed for each plate and the analyses were repeated for 5 days. Over 10' cattle tested, two false-positive results, one for sheep and one for poultry on the Bst plate, were obtained. The blank bovine and ovine samples gave IZ of 3.7 ± 0.7 and 3.3 ± 0.4 mm, higher than porcine (1.8 ± 1.4 mm) and poultry (2.0 ± 1.4 mm) blank samples. All blank samples gave non-specific inhibition on the plate Bst, higher than for

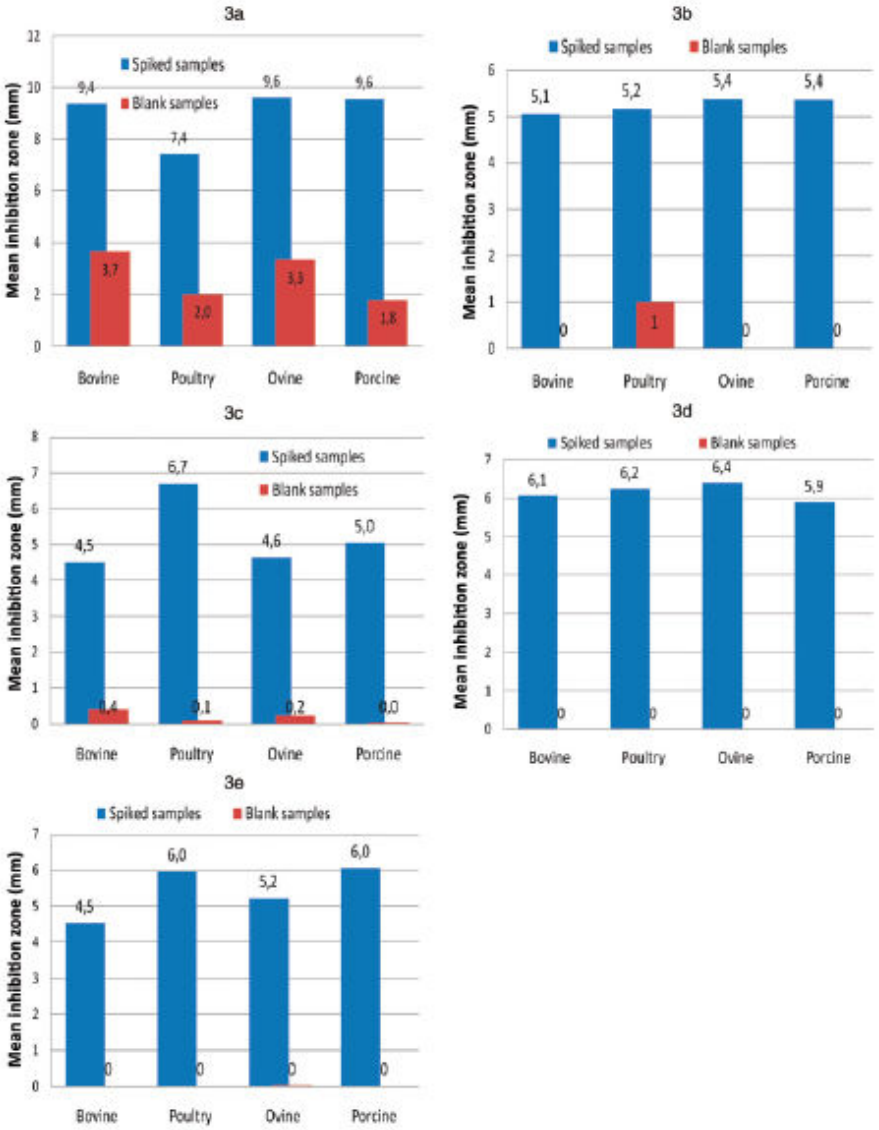


Figure 3. Results of the applicability study of the STAR protocol to muscles from different animal species, on the five different plates. The results of the applicability study for the five plates are represented: (a) penicillin G at 25 µg kg^{-1} on plate Bst, (b) doxycycline at 100 µg kg^{-1} on Plate Bc6, (c) erythromycin at 400 µg kg^{-1} on Plate Kv8, (d) enrofloxacin of 200 µg kg^{-1} on Plate Ec8, and (e) gentamicin at 6000 µg kg^{-1} on Plate Bs8. The *x*-axis shows the four different animal species tested are represented. The *y*-axis shows the average inhibition zone (mm) on the five different plates obtained during the applicability study for bovine, ovine and poultry muscles; and during the determination of CCβ for porcine muscle are represented for blank samples and spiked samples.

the other plates. That is why the positivity threshold of the plate Bst was set at 4 mm, while the positivity threshold for the other plates was set at 2 mm (Fuselier et al. 2000). No false-positive results were obtained on the plate, whatever the species, on the four other plates.

In conclusion, the STAR protocol is applicable for the screening of antibiotic residues in the muscle of the four major species: pig, cattle, sheep, poultry and by extension the muscle of minor species, whatever is the plate.

Table 5. Comparison of the mean inhibition zones (IZ) and the standard deviations (SD) calculated during the determination of the CCβ and during the ruggedness study.

Antibiotic family	Antibiotic	Tested concentration ($\mu g\,kg^{-1}$)	Plate	Mean IZ[a] (mm) ± SD CCβ	Mean IZ[a] (mm) ± SD ruggedness
Penicillins	Penicillin G	25	Bst	9.7 ± 0.2	8.5 ± 1.2
Sulfonamides	Sulfadimethoxine (SDMX)	300	Bst	8.8 ± 1.3	5.3 ± 1.6
Tetracyclines	Doxycycline	100	Bc6	5.4 ± 0.5	4.3 ± 0.6
Macrolides	Erythromycin	400	Kv8	5.1 ± 1.1	4.0 ± 1.0
Quinolones	Enrofloxacin	200	Ec8	6.2 ± 1.0	6.7 ± 1.2
Aminoglycosides	GTM	6000	Bs8	6.0 ± 0.7	6.2 ± 0.5

Notes: [a]IZ, inhibition zone of the specific plate, mean of the values of the four sets (two technicians).
SD, standard deviation.

Ruggedness study

During the determination of specificity and detection capabilities of the 16 antimicrobials, lots of different media (different preparation and validity date), different preparations of bacteria, and different batches of antibiotic standards were used over a period of 9 months of testing. Looking at the quantitative results (IZ) obtained by the two technicians on muscle samples, the variability (standard deviations) was rather low, similar to or lower than the variability that is fixed for the positive controls of the method. During this period, the positive controls were included in the intervals set in the STAR protocol. This is the first track to conclude that the STAR method is robust.

The average IZ and the standard deviations obtained during the ruggedness study and during the first part of the validation (determination of CCβ) were compared and are presented in Table 5.

When the standard deviation in ruggedness conditions (SDi) is significantly larger than the standard deviation of the method in terms of reproducibility, the conclusion is predictable that all factors taken together have an impact on the results, even if no single factor has significant influence. In this case, the method is not sufficiently robust regarding the range of tested variations. From a quantitative point of view, we did not observe significant differences between the SD in ruggedness conditions and the SD during the first part of the validation for the six tested antibiotics. Therefore, the STAR protocol seemed to be robust for the five plates regarding the range of tested variations.

If we look at the mean IZ obtained at each day (run) during the ruggedness study to the mean IZ obtained during the determination of detection capabilities, it seemed that the sensitivity is very near, even if the average IZ were often lower during the ruggedness study. In fact, the average IZ of SDMX on plate Bst was significantly lower during the ruggedness study (Table 5). Therefore, some of the tested factors had an effect on the sensitivity. Sulphonamides tested concentrations often gave PIZ, which are sometimes quite tricky to read for someone not trained in the technique. On the hand, beta-lactams usually gave clear IZ. Therefore, slight variations, like in the ruggedness study, even for trained people, lead to greater effect on these PIZ than on a clear IZ. Therefore, it is logical that the Bst plate is less robust for the detection of SDMX than for penicillin G.

The ruggedness of the STAR method was evaluated using an experimental design. The influence of the four factors (concentration of bacteria, quantity of medium, incubation time and pre-incubation time) on the mean IZ (quantitative result), repeatability (coefficient of variation, CV%), and false-positive and false-negative rate (qualitative result), as well as the interactions between factors, were evaluated (Table 2). The responses for each plate for each run (day) were indicated (mean IZ, CV% on the IZ, and false-positive and false-negative rates when influence was observed). From these results, the exploitation of the experimental design was performed. The results are presented in Table 6 for the six antibiotics and the five plates.

In conclusion, whatever the plates, there was no significant effect of the analytical factors tested on the qualitative results of the STAR method (no influence on the false-positive and false-negative rates), except for plate Bst with SDMX (Table 6). When testing SDMX, on day 4 the four tested samples gave false-negative results (Table 7). Therefore, it has a strong impact on the combination of factors tested that day: A+, B+, C−, D−, i.e., increasing the concentration of bacteria, increasing the amount of medium, reducing the incubation period, and without pre-incubation. In the optimization of microbiological methods, it is well known that increasing the concentration of bacteria or increasing the amount of medium may decrease the sensitivity of the plate. Moreover, as seen with other plates, a decrease of the incubation period can indeed reduce the sensitivity. It is also logical that a lack of pre-incubation may decrease the sensitivity of the plate. It was observed in fact that an increase of factors A and B may increase the

Table 6. Exploitation of the factorial design: influence of the four experimental factors on the detection of six tested antibiotics on the five plates of the STAR protocol.

		Factor				Interaction			Mean
Bc6	*Response*	*A*[a]	*B*[b]	*C*[c]	*D*[d] = *ABC*	*AB* + *CD*	*AC* + *BD*	*BC* + *AD*	*I*
	Mean IZ	−0.12	−0.01	−0.15	0.05	−0.13	0.01	−0.16	4.32
	CV (%)	−2.03	−3.08	5.23	2.75	2.91	−3.09	−1.88	11.21
Kv8	*Response*	*A*	*B*	*C*	*D* = *ABC*	*AB* + *CD*	*AC* + *BD*	*BC* + *AD*	*I*
	Mean IZ	−0.07	0.00	0.50[f]	0.28	0.12	−0.41	0.22	3.97
	CV (%)	−1.27	4.04	0.58	5.43	0.07	2.69	−1.93	16.07
Ec8	*Response*	*A*	*B*	*C*	*D* = *ABC*	*AB* + *CD*	*AC* + *BD*	*BC* + *AD*	*I*
	Mean IZ	−0.25	−0.58	−0.22	0.77	−0.45	−0.03	−0.33	6.72
	CV (%)	0.38	0.97	−0.07	−1.14	−0.47	−0.08	1.88	4.82
Bst	*Response*	*A*	*B*	*C*	*D* = *ABC*	*AB* + *CD*	*AC* + *BD*	*BC* + *AD*	*I*
SDMX	Mean IZ	0.34	−0.38	0.17	0.92	−0.30	−0.03	0.91	5.30
	CV	−0.26	0.15	−2.72	−0.01	0.25	−0.76	1.70	4.70
	False+	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	False −	0.50	0.50	−0.50	−0.50	0.50	−0.50	−0.50	0.50
Penicillin	Mean IZ	0.25	−0.20	−0.44	0.02	0.59	−0.42	0.03	8.55
	CV	−1.44	1.61	−2.91	1.74	−2.60	3.17	0.12	9.18
	False+	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	False−	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bs8	*Response*	*A*	*B*	*C*	*D* = *ABC*	*AB* + *CD*	*AC* + *BD*	*BC* + *AD*	*I*
	Mean IZ	−0.11	0.00	0.03	0.02	−0.11	−0.23	0.00	6.16
	CV (%)	0.67	−0.84	0.65	0.53	0.01	0.14	0.81	6.07

Notes: [a]Concentration of bacteria.
[b]Medium quantity.
[c]Incubation time.
[d]Pre-incubation time.
[f]Increasing the incubation time (C) for plate Kv8 increased the sensitivity of the plate a little (13%).
IZ, inhibition zone 5 mm; CV, coefficient of variation of the mean IZ (%); False+, false-positive rate; False−, false-negative rate. The mean false-negative rate is equal to 0.50, which means 50% of false-negative results.

Table 7. Ruggedness study for plate Bst: design matrix and experimental design calculation.

Run	Mean IZ SDMX	CV	False+	False−	Mean IZ Penicillin	CV	False+	False−
1	4.8	17.5	0	0	9.1	10.9	0	0
2	8.0	18.0	0	0	9.3	10.3	0	0
3	4.7	13.9	0	0	7.5	22.5	0	0
4	3.0	15.4	0	100[a]	10.0	4.6	0	0
5	5.3	10.2	0	0	9.1	2.0	0	0
6	4.6	7.6	0	0	7.5	7.1	0	0
7	5.1	13.4	0	0	7.5	7.1	0	0
8	6.9	11.8	0	0	8.4	8.9	0	0

Notes: Run 1 = day 1.
[a]A strong impact of the combination of factors tested on day 4 was observed on false-negative rate: A+, B+, C−, D−, i.e., increasing the concentration of bacteria, increasing the amount of medium, reducing the incubation period, and without pre-incubation.
IZ, inhibition zone (mm); False+, false-positive rate; False −, false-negative rate; CV, coefficient of variation (%) = (mean IZ)/(SD IZ)*100; SD, standard deviation.

false-negative rate, while an increase of the incubation time and pre-incubation of 1 h may increase the sensitivity. It is noticeable that 1 h of pre-incubation at room temperature increased the mean IZ of SDMX by 17%.

Regarding the classical variability of results of microbiological plate tests, none of the factors or combination of factors had a significant effect on the mean IZ (quantitative result) for any of the tested antibiotics on the five plates. Finally, no significant effect was observed on the repeatability of the method. Moreover, slight changes in the CV never had an effect on the qualitative results of the test for the tested antibiotics.

Therefore, the STAR protocol is a robust method for the detection of antibiotic residues in muscle. Some recommendations are given in the STAR protocol concerning the different incubation periods for the five plates, because incubation time is already known to be a critical parameter. In the range recommended in the protocol, the method is robust. It should be noticed that increasing the incubation time for plate Kv8 increased the sensitivity of the plate of 13% (Table 6). This is the reason why a longer incubation time (at least 24 h) compared with the other plates is recommended in the STAR protocol. However, even at the lowest incubation time (21 h), the results were satisfactory.

Finally, it was demonstrated that a pre-incubation of 1 h at room temperature could have a positive effect on the sensitivity of the test (plates Ec8, Bst (SDMX), Kv8) or no effect (plates Bc6, Bs8, Bst (penicillin G)) compared with no pre-incubation. However, pre-incubation is not a critical parameter because if people do not apply a pre-incubation, the results would be satisfactory, and if pre-incubation is performed, the results would be equal or better. A recommendation of a pre-incubation of 1 h at room temperature should be added in the next version of the STAR protocol. Moreover, on every day of the analysis, specific positive control antibiotic paper discs are placed on each plate. The results of these positive controls should be included in the range given in the STAR protocol. If it is the case, the results are valid.

Discussion

At this time, very few laboratories have tried to validate their screening methods (microbiological or immunological methods) according to Decision 2002/657/EC (European Commission 2002) for two main reasons. Firstly, Decision 2002/657/EC gave very little information and no technical recommendations for the implementation of the validation of a screening method. Secondly, especially for the validation of microbiological plate tests, the validation requires a much work, is time consuming and is quite expensive. In our laboratory, which is the Community Reference Laboratory (CRL) for antibiotic residues, the validation of screening methods has been a subject of interest for many years. At this time, the validation of immunological tests and microbiological tests (tube tests and plate test) were performed in our laboratory according to European Decision 2002/657/EC (Gaudin et al. 2004, 2007; Gaudin, Hedou, Rault 2009; Gaudin, Hedou, and Verdon 2009). The former 'detection limit' was replaced by a new performance characteristic called 'detection capability CCβ'. To our knowledge, the other validation studies of microbiological plate tests according to the European Decision (European Commission 2002) were from a National Reference Laboratory in the Netherlands (Pikkemaat et al. 2007, 2008; Pikkemaat, Oostra-van Dijk et al. 2009).

Regarding the time needed for this validation study, it would have been impossible to validate the STAR protocol for all the antibiotics having an MRL in muscle matrix (more than 50 antibiotics). The validation work would be too long, expensive and laborious. Therefore, the recommendation to use a list of representative antibiotics was a very good compromise and allowed to reduce drastically the scope of the validation. It has been proposed since the first version of the European guideline for the validation of screening method in 2005 to validate wide range tests only for a list of representative antibiotics. It was adopted in the final version of the guideline (Anon 2010). The choice of the representative antibiotics is not fixed and is dependent on different factors: the activity patterns of different antibiotics in one family, the matrix, the use of antibiotics in one specific country, and the assumed sensitivity of the method towards some antibiotics. The first step is to conduct a preliminary study that should allow one to determine a common pattern of activity for one family or at least several substances of the family on a specific class of bacteria. Therefore, one compound could be chosen to be representative of the other substances of the family in terms of the activity profile on bacteria. Then different antibiotics should be chosen, e.g., for milk or muscle matrices. In fact, the antibiotics used for intra-mammary treatment or for oral use could be different. Therefore, the interest of one antibiotic only used as an intra-mammary treatment is very limited for validation in muscle. Furthermore, if some antimicrobials are not used or not registered in some countries, there is no interest to validate for this compound if the method is intended for national control. Finally, the selected analytes are dependent of the analytical method. Therefore, if the method to validate clearly badly detect one antimicrobial, it is needless to determine its detection capability because this antibiotic would not be included in the scope of the method.

A similar proposition for validating a list of representative antibiotics was made by Pikkemaat, Oostra-van Dijk et al. (2009). After the determination of the activity profiles of 36 antibiotics for the NAT-post-screening test for the detection of antibiotic residues in kidney, the authors suggested that the validation could be performed on a list of representative antibiotics to reduce the scope of validation for routine field laboratories, for example.

One list of representative antibiotic was initially included as an example in the guideline which was extracted from the validation of the STAR protocol in milk (Gaudin et al. 2004). In fact, activity patterns were determined for 66 antimicrobials having an MRL

in milk. The conclusion was that several antibiotics in one family could be gathered into one group because they showed similar activity profiles (the same specific plate, a similar sensitivity). Each family could be divided into two or more groups. Finally, one or two antibiotics per family were chosen because their activity patterns were representative of one or two groups of antimicrobials in the same family. Compared with this list of representative antibiotics, some molecules were replaced by others during the validation study of the STAR protocol for the following reasons.

Some antibiotics are only used as intra-mammary treatment (e.g., cefalonium was replaced by ceftiofur because ceftiofur is widely used in cattle and swine to treat respiratory diseases, while cefalonium is only administered intra-mammarily). Moreover, it was determined during the validation of the STAR protocol in milk (Gaudin et al. 2004) that the least detected cephalosporins were cefquinome (MRL = 50 $\mu g l^{-1}$) and cefalexin (MRL = 200 $\mu g l^{-1}$). Oppositely, the best detected cephalosporins were cefazolin (Annex II for all tissues except milk) and cephapirin (MRL = 50 $\mu g l^{-1}$). Finally, ceftiofur (MRL = 1000 $\mu g l^{-1}$) was better detected than cefquinome, but less detected than cefazolin. Therefore, we have chosen to validate for ceftiofur and cefquinome because one was the least detected antibiotic of the family, with a low MRL (cefquinome), and the other (ceftiofur) has an intermediary detection with a high MRL. Regarding the activity profiles on the STAR protocol, it was assumed that if cefquinome is detected at 2000 $\mu g kg^{-1}$ (CCβ), cefazolin and cephapirin would be easily detected.

The antibiotic should be largely used in the country of implementation of the method (e.g., in France, sulfamethazine replaced sulfathiazole). Moreover, colistin, which belongs to the polymyxin family, was replaced by a second macrolide, tylosin, because macrolides are often used for animal treatment (cattle, swine, poultry).

During the development of the STAR protocol, it was already shown that the method had a very poor sensitivity for some antibiotics (e.g., flumequine, sulfaguanidine, colistin and spectinomycin), largely above the respective MRLs. Therefore, dihydrostreptomycin (DHS) replaced spectinomycin (level of detection in milk = 20,000–40,000 $\mu g l^{-1}$). During validation of the STAR protocol in milk (Gaudin et al. 2004), the level of detection of colistin was included between 200 and 2000 times its MRL (MRL = 50 $\mu g l^{-1}$ in milk and 150 $\mu g l^{-1}$ in muscle). Therefore, the STAR protocol is not fitted to the detection of colistin or of spectinomycin.

Some antibiotics are frequently detected in routine monitoring of samples: the tetracyclines, particularly oxytetracycline (OTC). Thus, tetracycline was replaced by oxytetracycline and doxycycline was added instead of flumequine. During the validation of the STAR protocol in milk, the limits of detection for OTC and TTC were both equal to 250 $\mu g l^{-1}$ (Gaudin et al. 2004). The detection limits of doxycyline (50 $\mu g l^{-1}$) and CTC (75 $\mu g l^{-1}$) were better. Therefore, the tetracycline family was divided into two groups from which OTC (the least detected) and doxycycline (the best detected) were chosen as representative compounds.

In a similar way, penicillin and cloxacillin were chosen as the representative compound for the penicillin family because penicillin G (LOD = 5 $\mu g l^{-1}$) was the best detected penicillin during the validation of the STAR protocol in milk, and cloxacillin (LOD = 60 $\mu g l^{-1}$) was the least detected penicillin (Gaudin et al. 2004). Moreover, they belong to two groups of MRL (penicillin G = 50 $\mu g kg^{-1}$ in muscle and cloxacillin = 300 $\mu g kg^{-1}$). Similarly, erythromycin (LOD = 30 $\mu g l^{-1}$) and tilmicosin (LOD = 50 $\mu g l^{-1}$) were the best detected macrolides in milk. On the contrary, tylosin (LOD = 200 $\mu g l^{-1}$) was the least detected and was representative of the group of spiramycin (LOD = 300 $\mu g l^{-1}$) and neospiramycin (LOD = 200 $\mu g l^{-1}$). Moreover, they also belongs to two groups of MRL (tylosin = 100 $\mu g kg^{-1}$ and erythromycin = 200 $\mu g kg^{-1}$). Concerning quinolones, enrofloxacin was considered as being representative of the quinolones family, except flumequine. In fact, flumequine is usually badly detected by many microbiological methods. When the STAR protocol was validated in milk, it was observed that enrofloxacin (LOD = 20 $\mu g l^{-1}$), ciprofloxacin (10 $\mu g l^{-1}$), marbofloxacin (30 $\mu g l^{-1}$) and danofloxacin (15 $\mu g l^{-1}$) had the same activity profiles, with similar sensitivities (Gaudin et al. 2004). The quinolone that was badly detected was flumequine (>600 $\mu g l^{-1}$). During validation in milk, sulfanilamide, sulfapyridine, sulfadoxine, sulfacetamide, and sulfaquanidine were the least sensitive sulphonamides (3.5–20 times the MRL (MRL = 100 $\mu g l^{-1}$) and the most sensitive sulphonamides were sulfaphenazole, sulfathiazole, and sulfachloropyridazine (MRL). The LODs of sulfadimethoxine (SDMX) (1.75 times the MRL) and sulfamethazine (2.5 times the MRL) were in between. The choice of sulphonamide was not based on the least detected compounds because it was too far from the MRL but was focused on two antibiotics that had intermediary sensitivities. Moreover, these two sulphonamides are commonly used for animal treatment. Concerning the aminoglycosides, gentamicin in milk was the best detected antibiotic (LOD = 300 $\mu g l^{-1}$) and spectinomycin was the least detected aminoglycoside (20,000–40,000 $\mu g l^{-1}$). The LOD for neomycin, kanamycin, streptomycin and DHS was 1000 $\mu g l^{-1}$. Therefore, DHS is representative of the three other aminoglycosides. Lincomycin was detected at 350 $\mu g l^{-1}$ and pirlimycin at 100 $\mu g l^{-1}$. Therefore, lincomycin was

the least detected and is representative of the detection of lincosamides.

The specificity of the STAR protocol for the detection of antimicrobial residues in porcine muscle was very satisfactory. A similar result was obtained by Pikkemaat et al. (2009) when implementing the STAR protocol on routine monitoring muscle samples in a comparative study. Only 1% of false-positive results were observed (6/591), on plate Bst, after the analyses of 591 routine monitoring samples, when the cut-off was set at 4 mm, which is the recommended cut-off in the STAR protocol for the plate Bst.

The results were obviously similar to the global orientation given in the STAR protocol: beta-lactams and sulphonamides were detected preferentially on plate Bst, tetracyclines on plate Bc6, quinolones on plate Ec8, and macrolides on plate Kv8. However, there were two exceptions. The aminoglycosides were detected on plate Bst with the concentrations tested for the determination of CCβ, whereas for higher concentrations than CCβ, they were detected on plate Bs8 (=Bs7.2). However, when the validation of the STAR protocol was performed in milk, Bs8 was the specific plate for the detection of aminoglycosides. Secondly, tylosin was detected preferentially on Bst instead of Kv8 at this concentration, which had already been observed during the validation of the STAR protocol in milk. Therefore, if a positive result is reported on plate Bst, the confirmation should be directed onto beta-lactams, sulphonamides, but also tylosin and aminoglycosides. A comparative study of three microbial screening tests including the STAR protocol applied to routine monitoring samples (Pikkemaat, Rapallini et al. 2009) has similarly shown that a macrolide (tulathromycin) was preferentially detected on plates Bc6 and Bst instead of plate Kv8. Therefore, if no tetracyclines could be confirmed in a positive sample on plate Bc6, the confirmation could be directed towards tulathromycin. During this comparative study, the STAR protocol was able to detect the four MRL samples that contained antibiotic concentrations higher than their respective MRLs. Three of them were tetracyclines, detected preferentially on plate Bc6, and the remaining residue was sulfadiazine at 172 $\mu g kg^{-1}$. No information was available for beta-lactams, macrolides and quinolones because none of these families was found in the routine monitoring samples. The NAT-screening test was able to detect four samples containing aminoglycosides, but not the STAR protocol because this method is not enough sensitive towards the aminoglycoside family. Furthermore, the NAT-screening test is applied to kidney, while the STAR protocol is recommended for muscle. Moreover, it is well known that aminoglycosides concentrated in kidney, while very low concentrations could be found in muscle. The muscle matrix is not a satisfactory matrix for the screening of aminoglycosides. The results of this comparative study are in accordance with our validation data.

The detection capabilities were determined for 16 antibiotics from different families in relation to their respective MRLs. The levels of detection of beta-lactams (penicillins and cephalosporins $\leq$MRL) were very satisfactory because even the least detected compound was detected at or below the MRL. Concerning tetracyclines, OTC (the least detected) could be detected at levels below 2.5 times the MRL and doxycycline (the best detected) at the MRL. The levels of detection of macrolides (2 MRL), quinolones ($\leq$2 MRL) and trimethoprim (2 MRL) were also satisfactory even for the least detected compound. The levels of detection of one representative sulphonamide SDMX ($\leq$3 MRL) were slightly higher than in milk, Finally, the sensitivity of the STAR protocol towards aminoglycosides and florfenicol was not satisfactory (>>MRL). Therefore, interest in plate Bs8 in the STAR protocol can be discussed. This plate should be improved for its sensitivity or should be replaced by another plate. Other possibility is to use a complementary method focused on aminoglycosides which could replace the use of this fifth plate.

The validation of the microbiological plate test is also a complex issue because of the solid matrices such as muscle and kidney. In fact, the ideal solution would be to produce incurred samples from animal treatment for all antibiotics having an MRL in the corresponding matrix and to validate all these antibiotics. However, this is extremely time consuming and expensive. Furthermore, it is really difficult to obtain exactly a target concentration after the treatment of animals. Therefore, there are different possibilities to implement the validation of a microbiological plate test.

The first option is to use spiked paper discs to determine the sensitivity of the method from antibiotic standard solutions (Koenen-Dierick et al. 1995; Calderon et al. 1996, Currie et al. 1998; Ferrini et al. 2006). The main advantages are that this solution is less expensive and quicker. The validation could be implemented for a wide list of antibiotic residues. However, matrix components could affect the detection capabilities of a method. Okerman, De Wasch et al. (1998) showed that only tetracyclines and quinolones were similarly detected with or without tissue. However, the detection of beta-lactams was better with antibiotic spiked discs without tissue than with tissues. For some antibiotics, the difference was only observed at low concentrations. Therefore, the use of antibiotic spiked discs is not totally satisfactory.

The second possibility is to use monitoring routine samples (Okerman, Van Hoof et al. 1998; Pikkemaat, Rapallini et al. 2009; Schneider et al. 2009). Routine samples from monitoring plans were analysed with different methods including the method to be validated

(including one or several physico-chemical confirmatory methods). The first disadvantage is that the number of samples containing residues is unknown. Moreover, very powerful confirmatory methods are needed to confirm all the samples, preferably even negative results at the screening step to check the false-negative rate of the screening test. Therefore, it is also costly and time consuming. Finally, only a narrow range of antibiotics (those which are most frequently used in the country) is encountered. Schneider et al. (2009) confirmed the presence of antibiotics in 29 samples, from which 23 belonged to the tetracycline family. This kind of validation led very to interesting information about the method, its practicability in routine conditions and some information on its performance, but the results are limited by the range of antibiotics and their concentrations.

Sometimes the validation study combined spiked discs and analyses of incurred materials, resulting from animal treatment (Myllyniemi et al. 1999, 2001). In this case, the authors used the activity patterns of antibiotic standard solutions to identify antibiotic residues in incurred materials. Furthermore, the STAR protocol was validated in our laboratory using spiked paper discs, plus analysing field routine samples. The interest of spiked discs is that many antibiotics at many concentrations could be tested, but the limitation is that no matrix effect could be observed. It is interesting to obtain preliminary data on the evaluation of a new method and activity profiles of many antibiotics. But we know that the matrix will affect the detection capabilities of the method. Therefore, field samples are very interesting because they are incurred materials and the matrix effect could be demonstrated. However, the limitation is that it is not known in advance how many positive samples are in the panel of field samples, which antibiotics would be detected and at which concentrations. Moreover, another limitation is that a little variety of antibiotic residues or families are usually found, depending on the matrix. Okerman et al. (2004) used artificially contaminated (spiked tissue fluid) as well as incurred samples for the comparative study of four screening methods for the detection of tetracyclines in muscle matrix. This approach is also interesting because the matrix effect could be observed both with spiked tissue fluid and incurred tissue.

The third possibility is the use of 'simulated tissues' as it is proposed in this paper. It seemed to be a good compromise between spiked discs and incurred tissues and the nearest preparation from intact muscles because the matrix effect can be evaluated. For the validation of a microbial screening assay, Pikkemaat et al. (2007) used two different approaches to produce fortified tissues. Firstly, spiked minced was heated and centrifuged to extract meat juice (supernatant) which was applied directly onto the plate. Secondly, incurred tissues were transformed into powder (using liquid nitrogen and blending the meat) and meat juice was extracted. Therefore, it was possible to obtain accurate concentrations in the samples by mixing incurred and blank materials. One conclusion was that 'the sensitivity of the assay for quinolones decreased two fold when matrix samples are analysed' by comparison with antibiotic standard solutions. In these conditions, the matrix effect was taken into account. Moreover, these two approaches are fit for this screening test because in a routine meat juice samples would also be analysed. However, that sample preparation differed a lot from the routine application of the STAR protocol which is based on pieces of raw muscle laid directly on the plates. Therefore, in our case, the 'simulated tissues' appeared to be as near as possible from real muscle samples. The binding of antibiotics to the minced tissue is possible during spiking. Pikkemaat et al. (2009) validated the NAT-post-screening test for the detection of antibiotic residues in kidney by determining the detection capabilities of 36 antibiotics in porcine as well as in homogenized fortified kidney samples (Pikkemaat, Oostra-van Dijk et al. 2009). After centrifugation, the supernatant was analysed. The same final treatment was applied to routine monitoring samples. This is another way of preparing simulated tissues.

Decision EC/2002/657 (European Commission 2002) has no recommendation concerning the kind of materials to be analysed (spiked or incurred matrix) during the validation. Therefore, all the previous validation studies could be in accordance with the European Decision if the way of determining performance characteristics was respected. The most important thing is to keep as close as possible to the materials analysed in routine use by the concerned method. Now the guideline for the validation of screening methods (Anon 2010) recommends the use of 'simulated tissues' (spiked tissues) when it concerns solid matrices and when it is impossible to obtain incurred tissues. Moreover, the validation could be restricted to a list of representative antibiotics. The validation, which was performed in the past on field samples, led to very interesting information concerning the method. However, this kind of validation does not fit in with Decision EC/2002/657 and with the recommendations of the validation guideline (Anon 2010). In fact, the number of antibiotics tested is always very restricted. Moreover, the number of samples to be analysed to determine the detection capabilities of antibiotics is never respected.

As a conclusion, there is no ideal way of validating microbiological plate tests that would not be time and money consuming. Obviously, validation of these kinds of methods needs a lot of effort from the laboratory that would validate the protocol for the first time (initial validation). In the guideline for the validation of screening methods (Anon 2010), one of

the recommendations is to reduce the work of validation when the method is transferred to another laboratory which would want to implement it in routine conditions. In fact, the number of samples to be analysed is reduced. The performance characteristics (e.g., detection capabilities) determined in the transfer laboratory have to be compared with those determined during the initial validation for a selected range of antibiotics. Then the participation to proficiency testing studies could complete the validation dossier.

Conclusions

The two objectives of this validation study were fulfilled. Firstly, the STAR protocol was validated for muscle by determining performance characteristics (specificity, detection capabilities CCβ for 16 representative antimicrobials, applicability, ruggedness), according to European Decision 2002/657/EC (European Commission 2002). The STAR protocol is applicable to the broad-spectrum detection of antibiotic residues in muscles of different animal species (pig, cattle, sheep, poultry). The method has good specificity (false-positive rate = 4%). The detection capabilities were determined for 16 antibiotics from different families in relation to their respective MRLs: beta-lactams (penicillins and cephalosporins $\leq$ MRL), tetracyclines ($\leq$MRL and $\leq$2.5 MRL), macrolides (2 MRL), quinolones ($\leq$2 MRL), some sulphonamides ($\leq$3 MRL), and trimethoprim (2 MRL). However, the sensitivity of the STAR protocol towards aminoglycosides and florfenicol was not satisfactory (>>MRL). Finally, this study has shown that the STAR method is a robust screening method, insensitive to reasonable variations analytical parameters such as the concentration of bacteria, the amount of medium, the incubation period, and pre-incubation or not.

Secondly, the applicability of the European guideline (Anon 2010) to the validation of a microbiological screening method and for muscle was demonstrated. The use of 'simulated tissue' appeared to be a very good compromise between antibiotic spiked discs and incurred samples. Moreover, the choice of a list of representative antimicrobials was also very interesting to reduce the scope of validation.

References

Anon. 2010. Guidelines for the validation of screening methods for residues of veterinary medicines (initial validation and transfer of method). Document of 21 January 2010.

Calderon V, Gonzalez J, Diez P, Berenguer JA. 1996. Evaluation of a multiple bioassay technique for determination of antibiotic residues in meat with standard solutions of antimicrobials. Food Addit Contam. 13:13–19.

Currie D, Lynas L, Kennedy DG, McCaughey WJ. 1998. Evaluation of a modified EC Four Plate Method to detect antimicrobial drugs. Food Addit Contam. 15:651–660.

European Commission. 1990. Council Regulation (EEC) No. 2377/90. 26 June 1990. Laying down a Community procedure for the establishment of maximum residue limits of veterinary medicinal products in foodstuffs of animal origin. Off J Eur Comm. L. 224:1–8.

European Commission. 1996. Council Directive (EEC) No. 96/23/EC of 29 April 1996 on measures to monitor certain substances and residues thereof in live animals and animal products. Off J Eur Comm. L. 125:10–32.

European Commission. 2002. Commission Decision 2002/657/EC of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and interpretation of results. Off J Eur Comm. L. 221:8–36.

European Commission. 2009. Council Regulation (EEC) No. 470/2009 of 6 May 2009 laying down Community procedures for the establishment of residue limits of pharmacologically active substances in foodstuffs of animal origin, repealing Council Regulation (EEC) No. 2377/90. Off J Eur Comm. L. 152:11–22.

Ferrini AM, Mannoni V, Aureli P. 2006. Combined plate microbial assay (CPMA): a 6-plate-method for simultaneous first and second level screening of antibacterial residues in meat. Food Addit Contam. 23:16–24.

Fuselier R, Cadieu N, Maris P. 2000. S.T.A.R.: Screening test for antibiotic residue in muscle. Results of a European collaborative study. In: van Ginkel LA, Ruiter A, editors. Proceedings of the Euroresidue IV Conference. Bilthoven (the Netherlands): National Institute of Public Health (RVM). p. 444.

Gaudin V, Hedou C, Rault A, Sanders P, Verdon E. 2009. Comparative study of 3 screening tests, Explorer® test, Premi®Test 2 microbiological tube tests and a multi-sulphonamides ELISA kit, for the detection of antimicrobial and sulphonamide residues in eggs. Food Addit Contam. 26(4):427–440.

Gaudin V, Hedou C, Sanders P. 2007. Validation of a Biacore method for screening eight sulfonamides in milk and porcine muscle tissues according to European Decision 2002/657/EC. J AOAC Int. 90(6):1706–1715.

Gaudin V, Hedou C, Verdon E. 2009. Validation of a wide-spectrum microbiological tube test, the EXPLORER® test, for the detection of antimicrobials in muscle from different animal species. Food Addit Contam. 26(8):1162–1171.

Gaudin V, Maris P, Fuselier R, Ribouchon JL, Cadieu N, Rault A. 2004. Validation of a microbiological method: the STAR protocol, a Five-Plate Test, for the screening of antibiotic residues in milk. Food Addit Contam. 21:422–433.

Koenen-Dierick K, Okerman L, De Zutter L, Degroodt JM, Van Hoof J, Srebnik S. 1995. A one-plate microbiological screening test for antibiotic residue testing in kidney tissue and meat: an alternative to the EEC four-plate method? Food Addit Contam. 12:77–82.

Myllyniemi AL, Nuotio L, Lindfors E, Rannikko R, Niemi A, Backmann C. 2001. A microbiological six-plate method for the identification of certain antibiotic groups in incurred kidney and muscle samples. Analyst. 126:641–646.

Myllyniemi AL, Rintala R, Backman C, Niemi A. 1999. Microbiological and chemical identification of antimicrobial drugs in kidney and muscle samples of bovine cattle and pigs. Food Addit Contam. 16:339–351.

Okerman L, Croubels S, Cherlet M, De Wasch K, De Backer P, Van Hoof J. 2004. Evaluation and establishing the performance of different screening tests for tetracycline residues in animal tissues. Food Addit Contam. 21:145–153.

Okerman L, De Wasch K, Van Hoof J. 1998. Detection of antibiotics in muscle tissue with microbiological inhibition tests: effects of the matrix. Analyst. 123:2361–2365.

Okerman L, Van Hoof J, Debeuckelaere W. 1998. Evaluation of the European four-plate test as a tool for screening antibiotic residues in meat samples from retail outlets. J AOAC Int. 81:51–56.

Pikkemaat MG, Mulder PPJ, Elferink JWA, De Cocq A, Nielen MWF, Van Egmond HJ. 2007. Improved microbial screening assay for the detection of quinolone residues in poultry and eggs. Food Addit Contam. 24(8):842–850.

Pikkemaat MG, Oostra-van Dijk S, Schouten J, Rapallini M, Kortenhoeven L, van Egmond HJ. 2009. Nouws antibiotic test: validation of a post-screening method for antibiotic residues in kidney. Food Control. 20:771–777.

Pikkemaat MG, Oostra-van Dijk S, Schouten J, Rapallini M, van Egmond HJ. 2008. A new microbial screening method for the detection of antimicrobial residues in slaughter animals: the Nouws antibiotic test (NAT-screening). Food Contr. 19:781–789.

Pikkemaat MG, Rapallini MLBA, Oostra-van Dijk S, Elferink JWA. 2009. Comparison of three microbial screening methods for antibiotics using routine monitoring samples. Analytica Chimica Acta. 637:298–304.

Renard L, Moulin G, Sanders P. 1992. Using experimental design to optimize a diffusion assay. J AOAC Int. 75(6):1045–1048.

Schneider MJ, Mastovska K, Lehotay Lightfield A, Kinsella Shultz C. 2009. Comparison of screening methods for antibiotics in beef kidney juice and serum. Analyt Chim Acta. 637:290–297.