SPECIAL GUEST EDITOR SECTION

Detection Methods and Performance Criteria for Genetically Modified Organisms

YVES BERTHEAU[1] and ANNICK DIOLEZ
Institut National de la Recherche Agronomique, PMDV/MDO, Route de Saint Cyr, 78026 Versailles Cedex, France
ANDRÉ KOBILINSKY
Institut National de la Recherche Agronomique, Biométrie, Domaine de Vilvert, 78352 Jour en Josas Cedex, France
KIMBERLY MAGIN[2]
Monsanto Co., 700 Chesterfield Pkwy N., St. Louis, MO 63198

Detection methods for genetically modified organisms (GMOs) are necessary for many applications, from seed purity assessment to compliance of food labeling in several countries. Numerous analytical methods are currently used or under development to support these needs. The currently used methods are bioassays and protein- and DNA-based detection protocols. To avoid discrepancy of results between such largely different methods and, for instance, the potential resulting legal actions, compatibility of the methods is urgently needed. Performance criteria of methods allow evaluation against a common standard. The more-common performance criteria for detection methods are precision, accuracy, sensitivity, and specificity, which together specifically address other terms used to describe the performance of a method, such as applicability, selectivity, calibration, trueness, precision, recovery, operating range, limit of quantitation, limit of detection, and ruggedness. Performance criteria should provide objective tools to accept or reject specific methods, to validate them, to ensure compatibility between validated methods, and be used on a routine basis to reject data outside an acceptable range of variability. When selecting a method of detection, it is also important to consider its applicability, its field of applications, and its limitations, by including factors such as its ability to detect the target analyte in a given matrix, the duration of the analyses, its cost effectiveness, and the necessary sample sizes for testing. Thus, the current GMO detection methods should be evaluated against a common set of performance criteria.

Genetically modified organisms (GMOs) were introduced in Europe at a time when consumer confidence was quite low, due to several independent and unrelated events. Chernobyl nuclear power crash, mad cow disease, *Listeria* or *Salmonella* food contamination, or the presence of *Legionella* bacteria into air-conditioning systems are some examples that crystallized consumers' fears. Most of these fears emphasized the need for accurate detection methods in fields such as food safety or plant protection. The increase of food quality standards also resulted in a large demand on origin and process traceabilities, controls, and food authentication. In the case of currently approved GMOs, this demand was not focused on a safety issue, but rather on the ability to provide new and innovative products to consumers and producers. With the tendency toward a general improvement of life quality, consumers now look for high-quality production standards and options in their food choices.

Detection of compounds or organisms is based partly or totally on physical, chemical, biological, and analytical methods. In the past few years, GMO detection has become a new application field of analytical methods, first dedicated to quality assessment such as seed purity and grain production, but now used to satisfy compulsory food labeling in a growing number of countries. By its application to the whole food supply chains in some countries, GMO detection constitutes a clear challenge for analysts as the first systematic application of several biological detection methods, such as nucleic acid-based amplification in a huge diversity of compounds or matrixes.

GM detection methods comprise a particular application of the more general field of analytical chemistry, which has benefitted from the practice of other analytical techniques. GMO detection methods continue to evolve through the preponderant accounting of methods and measurements based on performance criteria rather than prescriptive methods. In other words, the field of GM detection has now sufficiently matured to the point in which it is appropriate to establish and apply

Guest edited as a special report on "Method Development in Relation to Regulatory Requirements for Detection of GMOs in the Food Chain" by Elke Anklam and David Neumann.

Corresponding author's e-mail: bertheau@versailles.inra.fr.

[1] Task leader of the European standardization CEN/TC 275/WG 11 ad-hoc group for quantitative PCR methods.

[2] Task leader of the European standardization CEN/TC 275/WG 11 ad-hoc group for protein-based methods.

performance criteria. This review summarizes the value of these performance criteria approaches (1) as applied to GMO detection methods.

Importance of Performance-Based Methods and Measurements

Methods and measurements based on performance criteria constitute the background of a systematic approach that can satisfy the requirements of users, producers, and analysts, e.g., for validated methods. As shown in a national survey in 1998 in France, discrepancies of results in GMO detection were quite common some years ago. Part of these discrepancies clearly resulted from the lack of compatible and fully validated methods, i.e., of commonly accepted, validated through a collaborative trial, or applied in a frame of assurance quality and good laboratory practices (GLPs). Although many standardization bodies, such as the European Committee for Standardization (CEN) and International Organization for Standardization (ISO) working groups, request that methods of GMO detection be fully validated through collaborative studies, the number of GMOs to be detected, the number of matrixes to be analyzed, and the kinds of methods to be used (proteins vs nucleic acid-based methods, screening, insert specific, or identification by edge fragments, qualitative and quantitative, multiple vs simple sampling control plans) will not allow such extensive validations. Collaborative trial validation is an expensive, time-consuming, limited exercise.

Detection Methods and Performance Criteria

As a result of exponentially increasing requests for validated methods, the possibility for laboratories to develop, validate, and use their own in-house methods of analysis is well recognized. In-house or single laboratory validation is generally considered a scientifically and technically acceptable alternative to current international method validation practices. The validation of an internal method should be conducted on a more formal basis as it is already by a number of organizations in other fields of analysis (residues and veterinary products for instance). However, participation in collaborative trials still validates and supports comparability of results, variability between laboratories, and other pertinent performance criteria.

Current practices of analysis are also changing with the formal introduction of accreditation, proficiency testing, and defined internal quality control procedures into laboratories. The first step in a full-validation procedure should be to identify and document customer requirements and to identify the analytical problem, the analytical and economical possibilities, and other specific requirements on sampling, laboratory environment, external environment, personnel protection, etc. Thus, the validation plan should indicate the method criteria needed and address questions such as: (*1*) when is the method going to be used (official food control and in-house process control methods may have to fulfill different criteria such as precision and accuracy), (*2*) what type of answer is required—qualitative or quantitative, and (*3*) in what state is the analyte, i.e., bound, free, degraded, etc., fit for purpose?

Until now, the classical way to develop a standardized or reference method of analysis involved internal validation in one laboratory, followed by a collaborative or interlaboratory validation and subsequent acceptance when fulfilling requirements that were defined previously. However, this approach does not adequately address the growing number of proprietary and, thus, confidential methods, or the need to demonstrate that these methods perform as intended, producing accurate and reliable results. In addition, the growing accounting of performance criteria should provide analysts with a greater flexibility for validation of analytical methods, provided the method chosen meets certain predefined criteria.

The interest of such performance-based methods and measurements are their cost- and time-effectiveness, particularly at the validation level of the methods, their ability to facilitate the comparability and compatibility of the methods and, thus, of their data, and to provide objective criteria to accept or reject methods or data. However, given the present state-of-the-art of detection methods, a compromise is still needed between the previous and the more recent validation schemes.

GMO Detection and Performance Criteria

An approach based on minimum performance criteria to be met is also a growing concern at the standardization level of GMO detection. For instance, the French standard, released in December 2000, provides guidelines and general requirements in terms of performance criteria of the detection method, laboratory organization, GLP, and appropriate quality controls and results (2). It provides several goals for the analytical laboratories and gives a basis for the comparability of the methods and accreditation schemes. Such an approach has been recently taken into account at the international level, as exemplified by the CEN/TC 275/WG 11 general document and draft guidelines for the application of the criteria approach by a Codex Alimentarius committee on methods of analysis and sampling, both currently in preparation. However, in some particular fields, where standards are not available, laboratories should reach agreement with their clients on method performance specifications for each method and analysis, including reporting elements. From a practical point of view, standards based on method performance could be better established if a consensus could be reached between the experts for the values of the criteria, giving goals for future developments, and more rapidly reducing discrepancies of results by providing a frame for data reporting.

Thus, performance criteria-based methods and measurements are gaining increased popularity in the field of analytical methods, of which GMO detection is only a small part. The accounting of such an approach would drastically facilitate the standardization requested by consumers. The application of performance criteria could constitute a good example to the challenging field of GMO detection, providing a first step toward general guideline documents, requirements, and quality assurance.

Fields of Application of Performance Criteria in GMO Detection

Several unrelated methods are currently used to detect GMO, such as bioassays (sprays of herbicides on seedling), protein-based methods [enzymatic activity and immunological detection by strip tests and enzyme-linked immunosorbent (ELISA)], and nucleic acid-based methods (by direct hybridization in microarray or target or signal amplification by polymerase and ligase chain reactions (PCR and LCR), nucleic acid sequence-based amplification (NASBA), or self-replication (3 SR). Such methods can be either qualitative or quantitative, more or less specific of the insertion event (screening, insert-specific or event-specific, i.e., identifying the GMO unambiguously by a specific signature, such as the edge fragment of the GMO, provided that the homologous recombination is unavailable and the GMO does not result in gene stacking) or use a statistical approach of qualitative tests to make quantitative analyses [exhaustive dilution limits (3), or control plans such as that recommended by USDA-GIPSA (http://www.usda.gov/gipsa/biotech/starlink/starlink.htm)]. We are, thus, clearly facing an abundance of empirical methods whose compatibility on common plant tissues is still unknown. Performance criteria could provide a systematic frame to evaluate such methods.

Because most methods in use depend on the matrixes (raw to highly processed material) to be analyzed and the practices of the laboratories, requirements of customers or of countries, their compatibility can be achieved only if common performance criteria, as well as the domains of their application, have been previously defined.

Performance criteria can be defined in general guideline and requirement documents and then be used to define and develop analytical methods, validation (in-house or collaborative laboratory trials) and subsequent routine analyses from sample preparation. It includes design of control plans and their practice, definition of the laboratory sample and test portions and their preparation quality control, and proficiency testing. All methods should then meet the previously defined performance criteria, and, when appropriate, should comply with the regulatory requirements.

Of primary concern are also the performance criteria of the apparatus, the GLP and QA in metrology, and the maintenance and calibration of the apparatus.

Development of Methods and Performance Criteria

The development of methods and their internal validation, before any further collaborative or interlaboratory validation, is probably one of the less standardized aspects of the whole process toward routine use of detection methods. Although available for several years, the AOAC validation schemes (4) and literature issued from other analytical sectors (5) are rarely taken into account by molecular biology laboratories involved in GMO detection. For instance, the background of the staff can strongly influence the first internal validation as was observed during the French survey of 1998 of European laboratories. Clearly, although many molecular biologists performing PCR to detect GM traits were quite competent and skilled in the art, several controls and assurance quality parapets were forgotten in some laboratories. The training in basic research of most of these scientists did not include training in QA, classical method validation schemes of analytical chemistry, or other current practices of analysis, such as statistical assessment of repeatability and reproducibility.

From a practical point of view, several basic tools and methods are also missing, such as accurate methods to assay DNA in all or most situations (UV spectrophotometry, fluorimetry, image analysis of gel electrophoresis, exhaustive dilution limit). Several other standardized protocols, particularly statistical, to assess repeatability, reproducibility, uncertainty, limits of detection and quantitation, or rules to clearly specify the specificity of a detection method, would be also helpful to molecular biologists.

Validation of Methods and Performance Criteria

Validation of a method is defined as the process of determining the suitability of a measurement system for providing useful analytical data. Validation criteria should be the minimum necessary to ensure method performance for the intended purpose.

According to ISO/IEC Guides 25 and 5725, EMEA CPMP/ICH/281/95 guideline (6) and AOAC book on validation (4), there are a number of ways to validate analytical methods. Such an abundance of choice does not facilitate validation procedures whose statistical approaches do not take care of the goals and practicability searched for these methods.

It is the responsibility of the applicant to choose the validation procedure and protocol most suitable for the product. In practice, it is usually possible to design the experimental work so that the appropriate validation characteristics can be considered simultaneously to provide a sound, overall knowledge of the capabilities of the analytical procedure, for instance: specificity, linearity, range, accuracy, and precision. However, not all characteristics can be optimized; some performance criteria, such as accuracy, specificity, sensitivity, precision, or practicability, are more or less explicitly privileged. Such a situation clearly influences the validity of the validation procedures.

Collaborative Trials (Interlaboratory Comparisons)

Collaborative trials have long been used to establish the repeatability of methods and still constitute the most commonly practiced approach to method validation. They benefit from long experience, derived particularly from analytical chemistry, and of several important guidelines and requirements. General procedures defined through the ISO/IEC Guides 25 and 5725 or IUPAC/ISO/AOAC Protocol for the design, conduct and interpretation of method performance studies (7) constitute the bibles sustaining most of the interlaboratory validation studies (8, 9). However, several factors such as the minimum number of laboratories to be involved might evolve.

One may ask why the results of such expensive and time-consuming validations are largely underexploited. A

collaborative validation generally restricts the study to a statistical approach, discarding some participating laboratories (outliers) and to a final report of the observed variability. Such a validation does not try to identify the factors causing the observed variations. Unfortunately these expensive validations are not followed by feedback and corrective measure(s) to be proposed to the laboratories to better identify the cause of the variation and then to improve their practices. The introduction of experimental design (10, 11) into validation procedures in parallel with the current statistical validation approaches could largely improve the yield of collaborative validations and the consequent use of methods by considering the conditions of work of the laboratories by correcting the sources of variations. Further improvements of the collaborative validations, thus, are the current basis of standardization of methods. These improvements are particularly necessary for the emerging fields of analysis, such as the detection of GMO and generally all detection fields related to perceived and real fears.

Calibration Using References or Reference Materials

Validation of methods through the use of reference material is a very common practice for most methods of analysis. However, the availability of internal or certified reference material is a bottleneck for almost all methods, from their development to their routine use. The production and use of this reference material is itself placed under the auspices of performance criteria. Stability and continuity are some of the major factors affected by the definition of performance criteria. When necessary, these reference materials should be linked to an international standard, which is not currently the case for the GMO. Calibration is also a domain largely prone to errors and bias, particularly for quantitative methods. The compatibility of real-time PCR and statistical and accurate approaches, such as exhaustive dilution limits (3), is an open question, while the intrinsic variability of real-time PCR remains unknown, which impairs the ability to accurately calibrate the measures.

Large discrepancies of results can then be expected from such a situation, particularly when reference material is unavailable to assess the differences as is currently the case. As yet, no comparison of the several detection methods (e.g., screening vs construction-specific or insert-specific or protein- vs DNA-based detection methods) on the same material or on appropriate tissues originating from the same GMO has been undertaken. This validation plan supposes the availability of material, if possible common to the several methods in use (bioassays, protein- and nucleic acid-based methods) or with scope perfectly defined. In all these cases, the performance criteria can be defined only if clear correlation can be established between the several analytes, whose amount can vary between tissues, according to the genetic background and the physiological states [e.g., the different content of CryIA(b) protein of Bt176, Mon810 etc.].

In the present state of the GMO commercialization, plants are the only GMO concerned. Seeds are, thus, the material of choice to constitute the reference. What level of purity (95 or 100%), a characteristic similar to a performance criteria, should we request for such reference material? While some are asking for an expensive and quite biologically and economically inaccessible target, it is more realistic for the reference material to comprise commercial, certified seeds, the common source of worldwide plant production. Stability and continuity is a basic request for reference material whose practical conditions of production, and performance criteria, are still to be discussed. As an example of such a need of continuity, several papers indicated that the Starlink corn production might cease, which would suppress a source of standards while other observers think that Starlink-derived compounds will be present in the food supply chain for several years.

Pure DNA is an alternative to commercial seeds, which could resolve several aspects such as the stability and continuity of standards. Such material could also allow us to define more easily the optimized conditions of the tests to meet the performance criteria, such as the absolute detection and quantitation limits or the internal variability of PCR tests. Such a reference material is therefore requested by most laboratories. However, the ability of methods, using only DNA as reference material, to comply with regulatory aims such as the European requirement based on the ingredient (analytically translated as plant species) weight basis, is an open question, particularly if the relationships between DNA and the weight are not formally established. Indeed, the analyte content varies among tissues as observed for seeds and other tissues whose DNA content can vary under temperature conditions from 2 to 384C (12, 13). Silencing of transgenes in monocots is also a more common situation than previously expected (14), whereas weather conditions during growth, maternal background, or infections by plant pathogens can also alter the regulation of the inserted genes or seeds size (15). The resulting content of analytes (proteins or nucleic acids) could then greatly differ among reference materials and the conditions of raw material production (seeds) should be designed carefully. Thus, basic research programs on the effect of several factors on DNA and proteins content of seeds, the classical reference material, should be performed before any stable and accurate reference material is released.

Comparison of Results with Other Methods

Meta-analysis is a statistical procedure that integrates the results of several independent studies considered to be combinable. It provides a more objective appraisal of the evidence than traditional narrative reviews, a more precise estimate of a treatment effect, and may explain heterogeneity between the results of individual studies. However, as with all methods, it can be biased by the exclusion of relevant studies or inclusion of inadequate studies (16). Moreover, several statistical approaches (regression method, rank correlation) are available whose abundance does not help to refine the results. However, such an approach has been used, for instance, to determine several factors involved in QA (17). Although beginning to be used in the medical field (18), meta-analyses of previous validations tests are still unexplored in GMO detection to take into account and synthesize results systematically.

Systematic Assessment of Factors Influencing Results

This kind of method validation is highly dependent on the knowledge issued from laboratories and from the accumulated experience of the staff with practices and apparatus.

The growing interest in this approach is supported by the better consideration of the experimental design plans. However, the practical use of this validation approach is limited by the intellectual investment and time-consuming preparation involved. As yet, such an approach is restricted mainly to the development steps, the in-house validation, or the collaborative validations among a limited number of laboratories. It provides, probably better than proficiency testing, improved laboratory practices and might objectively refine performance criteria by the experience gained by the laboratories and through its systematic approaches. Experimental design, however, involves a good identification of the factors to be taken into account and, thus, the experience of the staff.

Assessment of Uncertainty of Results Based on Scientific Knowledge and Practical Experience

Uncertainty is a parameter associated with the result of measure, which characterizes the dispersion of values and is reasonably attributed to the method of measure. It is associated with a tolerance accepted for a method. Well-defined practical methodology is needed to develop meaningful data to assess uncertainty in GMO analysis, particularly at the low level. Measurement uncertainty should be estimated, if required, and be available. Until now, the acceptable tolerance for GMO detection by quantitative PCR and particularly real-time quantitative PCR still remains unknown, while several factors, such as the origin of the DNA polymerase, influence its accuracy (19).

Current validation procedures scarcely address compliance to explicitly predefined performance criteria but do address results of interlaboratory studies, which are only a part of the total picture. Experimental design, performance criteria, and the associated guidelines and requirements could identify the origins of variability to improve both methods and laboratory practices.

Routine Analyses

Routine analyses are easier to control for their accuracy as they are the first target of all controls applied in a laboratory, either by direct determination or by using values previously established from validation of the method. It must, however, be recognized that routine analysis is also sensitive to normal assay drift, which is known to occur with routine performance of any method. Given the inherent variability of a method in routine use, quality controls are useful to monitor ongoing conformity with the performance characteristics established during initial validation. However, routine results can also be used to refine the established performance criteria, provided analysis of routine data through experimental design is applied.

When samples are analyzed anonymously, proficiency testing is part of quality control. Analyses of fortified test portion and reference material (certified or not) provide an estimate of precision and bias of the analytical methods and of the assay drift of routine analyses. Again, the growing number of proprietary internal methods and kits in GMO detection emphasizes the need for performance criteria and related guidelines. Participation in proficiency testing can open the door to accreditation schemes. When a laboratories intends to use an unfamiliar method, such as detection kits, it is the responsibility of the laboratories to verify that it is competent to use the method. Each laboratory, experienced or novice, should demonstrate proficiency with each new method to be performed in an analytical setting, including appropriate quality controls necessary to assess the competence of the laboratory and the accuracy of results.

Performance criteria, then, should be more explicitly considered in all steps necessary to perform a given detection method. While quality control, GLP, and QA are necessary for proper method development and internal method validation, much improvement is needed overall, especially in the area of method development, which drastically lacks recognized and accepted performance criteria. Most development efforts appear quite crude and should be more formalized according to the currently available literature (4). Development efforts have not sufficiently matured at this time to readily accept adoption of performance criteria during the initial method development. Although in-house validation is clearly growing in terms of its contribution to a comprehensive method validation, interlaboratory or collaborative trials are necessary to define the variability acceptable from laboratory to laboratory and are a requirement for standardization of some reference methods.

Performance Criteria

Many of the performance criteria are well recognized and accepted by numerous internationally recognized organizations (ISO, EMEA, national standardization bodies, Codex Alimentarius, EURACHEM, etc.). Such criteria may appear commonplace in association with analytical methods, but a brief reminder of the definitions is perhaps beneficial. Probably one of the better sets of performance criteria and definitions has been provided by the Budapest Document of the Codex Alimentarius and documents from the recent meeting of March 2001. These criteria are as follows:

(**a**) *Applicability*.—The scope of application of the method should identify the matrix, analyte, or species being measured, its concentration range, and the type of study/monitoring effort for which the procedure, as judged from its performance characteristics, is suited. It should also describe the known limitations of the method.

(**b**) *Specificity or selectivity*.—The selectivity is the ability to discriminate between the analyte or species to be determined and other materials in the test sample.

(**c**) *Calibration*.—The calibration or standard curve is a graphic representation of the measuring signal (the response variable) as a function of the quantity of analyte or measurand.

(**d**) *Accuracy (also often called trueness)*.—Accuracy is the closeness of agreement between a test result and the accepted reference of true value of the property being measured, for example the true content of a specific analyte in a sample.

(**e**) *Precision*.—Precision is the closeness of agreement between independent test results obtained under stipulated conditions. Repeatability, intermediate precision, and reproducibility are often associated with this performance criterion.

(**f**) *Range (or dynamic range)*.—The interval of concentration within which the analytical procedure demonstrates a suitable level of precision and accuracy. The linearity is also often associated with this characteristic.

(**g**) *Limit of quantitation (LOQ)*.—The limit of quantitation of an analytical procedure is the lowest amount of concentration of analyte in a sample which can be determined quantitatively with an acceptable level of precision and accuracy.

(**h**) *Limit of detection (LOD)*.—The limit of detection is the smallest amount or concentration of analyte in the test sample that can be distinguished reliably, with stated significance, from the background or blank level.

(**i**) *Sensitivity*.—The sensitivity of a method is a measure of the magnitude of the response caused by a certain amount of analyte.

(**j**) *Ruggedness*.—The ruggedness of an analytical method is the resistance to change of an analytical method when minor deviations are made in the experimental conditions of the procedure.

(**k**) *Practicability*.—The ease of operations, in terms of sample throughput and costs, to achieve the required performance criteria and thereby meet the specified purpose.

This last performance criterion clearly outlines the impact of costs, a factor often forgotten for GMO detection. Such costs can be split into (*1*) investment costs as the development costs, investments in specific and material or personnel direct and indirect costs (training and protection of the personnel); and (*2*) routine analysis costs, which are also relevant to this practicability not only in terms of reagents, but also in terms of duration of analysis both for possible food immobilization and personnel costs, and in terms of the cost of the matrix to be analyzed itself (*see* the costs of breeder seeds, for instance).

Except the last one, all these performance criteria are taken into account into the current working document of the CEN/TC 275/WG 11. However, there are some inherent difficulties to meet these criteria in the evolving field of DNA-based detection. Values of performance criteria should be defined, provided that clear methods of calculation are defined. For instance, several ways are commonly used to define limits of detection and quantitation: visual evaluation, signal-to-noise ratio, standard deviation of the response and the slope, standard deviation of the blank, calibration curve, and finally, recommended data. In this case, the choice of calculation method should affect our ability to discern the compliance of analytical methods with the performance criteria values (21) and the possibility of inducing new sources of discrepancy between analytical laboratories.

Ways to Define Values of Performance Criteria

The first question about values of performance criteria might be as such: Do we have to define first values and to develop methods to meet them or, alternatively, do we have to accept the observed performance, for instance, during collaborative trials?

Although this question could be raised for each above-mentioned performance criterion, it is particularly acute in the case of the accepted or acceptable variability of quantitative measure. The systematic analysis of Horwitz (5) and subsequent works clearly show that the variability is correlated to detection and quantitation limits, increasing with the decreasing values of the limits. As the approved GMOs are not a safety concern under the current state of knowledge, it is appropriate to define LOD and LOQ for GMO detection methods that satisfy not only consumers and regulations, but also industrial constraints, and should restrict the sources of result discrepancies and potential legal actions. That is the way chosen by the French standard published in December 2000 (2). Such a pragmatic approach is also intended to decrease the technological race observed between service laboratories during a European survey, which were trying to increase the LOD and LOQ of their in-house developed methods as well as the test portion size in order to retrieve enough DNA for analyses.

Factors Affecting Values of Performance Criteria of GMO Detection Methods

Performance criteria can be applied to the whole process of detection. This section emphasizes the practical implications and strategies available throughout the whole detection process.

Sampling and Subsampling

Sampling is the critical starting point of all detection methods. Although numerous theoretical descriptions of appropriate sampling plans have been described, their practical application is still unresolved. Because the best sampling can be done only in moving matrixes, such as grains in elevators, the criteria of performance are rarely fully met. The sampling strategy is also dependent on the kind and costs of matrix to be analyzed, the cost of the analytical method per data point, and the risks accepted in a contract by buyer and seller. A single control plan protects mostly the buyer and creates disproportionate producer risk.

In some instances, the prices and/or scarcity of the matrix, such as breeder seeds or Elite seeds, clearly favor the multiple sampling plans (multistage plans), such as that recommended by USDA-GIPSA, which allows the use of fewer grains for similar detection and quantitation levels. Other interests of multiple sampling plans are their ability to turn a set of qualitative data into quantitative and to be independent of other sources of artifacts and variability, such as the empirical correlation used in the quantitative real-time PCR between the fluorescence and the initial number of DNA molecules.

Although the approach of multiple sampling plans is one of the most robust, it is not always applicable when, for instance, the unit price of the analytical method is too important: the price of PCR tests for processed foods with low contents of DNA prohibits multiple sampling plans and testing of multiple samples. The French standard retains a laboratory sample size of 10 000 grains or its mass equivalent for most matrixes for a 99.9% confidence interval. Such a large quantity of material induces drastic needs for new material that is easy to clean and the ability to grind the sample to <0.5 mm for homogenizing, as well as the subsequent reduction of the sample to a size appropriate for further fine grinding and sampling of the portions. Theoretically, to facilitate target extraction, the appropriate grinding of portions, particularly for raw material such as grains, should be as close as possible to that of the cell sizes. Generally, economical constraints strongly influence the values of performance criteria through the choice of sampling methods.

The definition of the size portion is often debated. Some laboratories advocate increasing the size of the test portion until enough DNA is retrieved to satisfy the LOD and LOQ. As safety has been established for the approved GMO, the French standardization commission, mirror of the CEN/TC 275/WG 11 working group, decided to limit the size of the test portions of most matrixes to 1 g. This decision induced the modification of numerous protocols and created an inability of most commercial extraction kits to work with such a size.

Extraction

As previously outlined, the efficiency of the protein or DNA extraction is influenced by the size of the ground particles submitted for analysis. This mass recovery is not correlated to quality of the extracted target, and more often than expected, inhibitors of immunological or DNA-based methods reduce the LOD and LOQ. Extraction is, thus, usually a multiple step process whose result should be carefully checked either by using similar matrix for protein-based methods or by using appropriate internal controls, such as those spiked for DNA-based methods. The matrix effect is part of the domain of application required by European standards for all detection methods proposed for standardization.

As yet, there are no performance criteria established to assess the efficiency of recovery and the quality of extracted DNA. Methods based on spiked DNA or proteins are not really representative of the state of the analyte in the matrix. Moreover, the LOD and LOQ specified often involve DNA amplification to be performed on a minimum quantity of DNA. Further, there are no satisfactory methods for quantitation of DNA. UV spectrophotometry is applicable only for a rough assessment of raw material such as grains. This technique largely overestimates the quantity of DNA by the hyperchromicity of degraded DNA and by taking RNA into account. Moreover, it provides crude and inaccurate information on the quality of DNA (sizes of the extracted fragments, kind and quantities of reaction inhibitors). The fluorimetric approach with DNA fixing molecules specific of dsDNA is efficient for the low content of molecules, but does not provide information about DNA quality. Although probably the most efficient way to quantitate extracted DNA, exhaustive dilution limit can be used, although with difficulty in routine analysis. Finally, the several DNA assay methods are not easily correlated.

An interesting way of DNA quantitation and quality assessment would be PCR tests on universal sequences, provided the number of copies is constant among the organisms. However, such a user-friendly approach for quantitation and quality of DNA still remains to be developed. In conclusion, several practical factors affect our ability to assess compliance of detection methods with performance criteria as they also affect the observed variability of interlaboratory studies.

Conclusions

The performance criteria approach is a basis to establish and use detection methods that should facilitate all the steps from development to routine GMO analyses. As the number of methods, matrixes, and GMO increases exponentially, the in-house or single-laboratory validation of proprietary or public methods is also largely growing. More than for the interlaboratory (collaborative) validation studies, such kind of validation needs consensual and well-defined performance criteria as a common basis for all laboratories. Performance criteria are used more or less implicitly by collaborative trials. Essentially, the growing consideration of performance criteria-based methods and measurements results from a need of transparency by clearly providing, before any expensive and time-consuming work, the criteria of acceptance or rejection of methods and/or data.

In place of interlaboratory studies, noting some performance of the techniques, realistic, fixed values of performance criteria would facilitate the compatibility among methods and provide a common ground for new objectives and research fields. In such studies, experimental design should add values by determining the sources of variability to be corrected. Fixing values for performance criteria is a compromise between technical requirements, limitations, and compliance with regulations, as is the case for GMO detection, and practical considerations such as duration and cost of analyses.

Although the use of performance criteria is playing an increased role, it is highly probable that performance criteria and its correlated guidelines and requirements will meet obstacles: awareness, familiarity, agreement, self-efficacy, outcome expectancy, ability to overcome the inertia of previous practice, and finally the absence of external barriers to perform recommendations. Again, a compromise will have to be found (22).

References

(1) Thompson, M., Ellison, S.L., & Wood, R. (1999) Harmonized Guidelines for Single-Laboratory Validation of Methods of Analysis, Technical Report IUPAC Symposium on Harmonization of Quality Assurance Systems for Analyti-

cal Laboratories, Budapest, Hungary, November 4–5, http://www.iupac.org/

(2) AFNOR (2000) *T1 Foodstuffs, T2 Detection and quantification of genetically modified plant organisms and derived products, T3. Part 1: Guidelines and Requirements,* XP V 03-020-1 Standard, Association Francaise de Normalisation, Saint Denis La Plaine Cedex, France, Contact: catherine.guisse@afnor.fr

(3) Rodrigo, A.G., Goracke, P.C., Rowhanian, K., & Mullins, J.I. (1997) *AIDS Res. Human Retrovirus* **13**, 737–742

(4) Wernimont, G.T. (1985) *Use of Statistics to Develop and Evaluate Analytical Methods*, AOAC, Arlington, VA

(5) Schofield, T.L. (2000) in *Encyclopedia of Biopharmaceutical Statistics*, Shein-Chung Chow (Ed.), Marcel Dekker, Inc., New York & Basel, p. 535

(6) EMEA (1995) CPMP/ICH/281/95, European Agency for the Evaluation of Medicinal Products, http://www.emea.eu.int/htms/ich/quality/ichfin.htm

(7) Horwitz, W. (1995) *Pure Appl. Chem.* **67**, 331–343

(8) Lipp, M., & Anklam, E. (2000) *J. AOAC Int.* **83**, 919–927

(9) Lipp, M., Bluth, A., Eyquem, F., Kruse, L., Schimmel, H., Van den Eede, G., & Anklam, E. (2001) *Eur. Food Res. Technol.* **212**, 497–504

(10) ASU (1997) *Plans d'expérience: Applications à l'entreprise*, J.J. Dresbeke & G. Saporta (Eds), Technip Editions, Paris, France

(11) John, P.W.M. (1971) *Statistical Designs and Analysis of Experiments*, Macmillan, New York, NY

(12) Schweizer, L., Yerk-Davis, G.L., Phillips, R.L., Srienc, F., & Jones, R.J. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 7070–7074

(13) Kowles, R.V., & Phillips, R.L. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7010–7014

(14) Iyer, L.M., Kumpatla, S.P., Chandrasekharan, M.B., & Hall, T.C. (2000) *Plant Mol. Biol.* **43**, 323–346

(15) Lemontey, C., Mousset-Déclas, C., Munier-Jolain, N., & Boutin, J.-P. (2000) *J. Exp. Bot.* **51**, 167–175

(16) Sterne, J.A., Gavaghan, D., & Egger, M. (2000) *J. Clin. Epidemiol.* **53**, 1119–1129

(17) Horwitz, W., Kamps, L.R., & Boyer, K.W. (1980) *J. Assoc. Off. Anal. Chem.* **63**, 1344–1354

(18) Owens, D.K., Holodniy, M., McDonald, T.W., Scott, J., & Sonnad, S. (1996) *Ann. Intern. Med.* **124**, 803–815

(19) Kreuzer, K.A., Bohn, A., Lass, U., Peters, U.R., & Schmidt, C.A. (2000) *Mol. Cell Probes* **14**, 57–60

(20) Codex Committee on Methods of Analysis and Sampling (1998) *In-House Method Validation*, Codex Alimentarius Commission, CX/MAS 98/8, Budapest, Hungary, November 23–27, http://www.fao.org/

(21) Rodbard, D., Munson, P.J., & De Lean, A. (1978) *Improved Curve-Fitting, Parallelism Testing, Characterization of Sensitivity and Specificity, Validation and Optimization for Radio-Ligand Assays*, IAEA-SM-220/58, Vienna, Austria, pp 469–503

(22) Cabana, M.D., Rand, C.S., Powe, N.R., Wu, A.W., Wilson, M.H., Abboud, P.A., & Rubin, H.R. (1999) *JAMA* **282**, 1458–1465