codex alimentarius commission

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

WORLD HEALTH ORGANIZATION

JOINT OFFICE: Viale delle Terme di Caracalla 00100 ROME Tel: 39 06 57051 www.codexalimentarius.net Email: codex@fao.org Facsimile: 39 06 5705 4593

Agenda Item 9 **CX/MAS 04/10**

JOINT FAO/WHO FOOD STANDARDS PROGRAMME
CODEX COMMITTEE ON METHODS OF ANALYSIS AND SAMPLING
Twenty-fifth Session, Budapest, Hungary, 8 – 12 March 2004

CRITERIA FOR THE METHODS FOR THE DETECTION AND IDENTIFICATION OF FOODS DERIVED FROM BIOTECHNOLOGY GENERAL APPROACH AND CRITERIA FOR THE METHODS

BACKGROUND

At the Twenty-fourth Session of the Codex Committee on Methods of Analysis and Sampling, papers giving the methods that had been collated by the *ad hoc* Intergovernmental Task Force on Food Derived from Biotechnology (see CX/MAS 02/8) and outlining general considerations of methods of analysis for the detection and identification of foods derived from biotechnology (see CX/MAS 02/9) were discussed. It was noted that the presence of genetically modified organisms or their derivatives could be assessed by the detection of either DNA sequences present as a result of recombination or the protein coded by the inserted gene. It was pointed out that protein-based methods were cheap, offered high selectivity and sensitivity but that since proteins were denatured during processing these techniques were most suitable for the analysis of raw materials and were not generally applicable to highly processed foods. It was also noted that these methods cannot be used when no protein is expressed in the food.

It was also noted that presence of a genetically modified organism or its derivatives could be established by the detection of either DNA sequences present as a result of recombination or the protein coded by the inserted gene. Methods of detection of DNA markers based on the polymerase chain reaction (PCR) have been used in a variety of food analyses and widely used for detection of GM derivatives in food for many years, and modifications of the PCR method were also widely used. A typical method involved several steps such as extraction and purification, amplification by PCR and detection/quantification. Specific questions arising in the area of proficiency testing, use of performance criteria and the necessity of quantification of the threshold since the results of investigations showed the difficulties in measuring low levels of GM material in processed foods were also discussed. Methods described in the collated documents could only be used successfully if all information about the sequence and standard reference materials were available.

GENERAL CRITERIA

In view of the absence of precise provisions for GMOs and of difficulties with the practical application of methodology in this area, the Committee proposed to develop recommendations with respect to criteria for methods of analysis and for quality control measures that should be introduced in laboratories offering GM analyses. It was agreed that a Working Group led by Germany and the United Kingdom would update and further develop the paper for this session and prepare recommendations for quality control measures in laboratories and criteria for methods of analysis for the Twenty-fifth Session of CCMAS. The following countries and organisations expressed their willingness to participate in this work: Argentina, Australia, Brazil, Canada, Egypt, France, Iran, Ireland, Italy, Japan, Malaysia, Netherlands, Philippines, United States, European Commission, AOAC, AOCS, EUROPABIO, and ISO.

These measures are given as Guidelines in the Appendix to this paper.

RECOMMENDATIONS

It is recommended that the draft Guidelines be discussed at the Twenty-fifth Session of CCMAS. If there is sufficient consensus, then the approaches described should be further refined and developed by a Drafting Group and then sent to governments for comment.

APPENDIX I: GUIDELINES FOR THE VALIDATION AND QUALITY CONTROL REQUIREMENTS FOR GMO ANALYSES

INTRODUCTION

Method Criteria

The Codex Alimentarius Commission places an emphasis on the acceptance of methods of analysis which have been "fully validated" through a collaborative trial conforming to an internationally accepted protocol. In a number of sectors, including the GMO sector, there are few methods of analysis which have been fully validated. As a result, Codex is also endorsing by reference single-laboratory validation protocols. In the GMO area there may be pressure to adopt a formal single-laboratory validation as an interim measure in the absence of collaborative trial data.

Many methods are currently being developed for GMO detection, identification and quantification. Before they are accepted for use by Codex they must be validated to ensure that they are fit-for-purpose.

The conventional criteria that have been adopted by Codex for the evaluation of methods of analysis are:-

- accuracy
- applicability (matrix, concentration range and preference given to 'general' methods)
- detection limit
- determination limit
- precision; repeatability intra-laboratory (within laboratory), reproducibility inter-laboratory (within laboratory and between laboratories), but generated from collaborative trial data rather than measurement uncertainty considerations
- recovery
- selectivity
- sensitivity
- linearity

These Guidelines address these requirements in the GMO sector, and anticipates that is likely that these will have to be further expanded by other items such as:-

- amplicon length
- whether the method is instrument specific
- whether there are differences between qualitative and quantitative PCR-based detection methods
- whether there are differences between PCR-based and immunological methods concerning validation criteria
- whether single- or multi-plex PCR amplifications are undertaken
- maximum number of PCR cycles

The method validation process accepted by Codex includes the definition of the requirements for the method, testing that the method meets these requirements when carried out, for instance, by different laboratories in different countries, and documentation of the method performance and measurement uncertainty.

Criteria approach

Codex Alimentarius Commission has accepted the "criteria approach" for methods of analysis. This approach does not extend to Codex Type I empirical/defining, procedures. It is necessary to ensure that this approach is incorporated into Codex guidelines on the validation of GMO methods of analysis unless it is explicitly stated that all GMO methods of analysis are empirical, both theoretically as well as in practice.

Laboratory Quality

The Codex Alimentarius Commission has adopted guidelines for the "quality" of laboratories involved in the import and export of foods. These quality characteristics are based on accreditation to ISO/IEC Standard 17025, proficiency testing and internal quality control as well as the use of methods of analysis validated according to Codex requirements. These overarching guidelines provide information to and dictate requirements for laboratories working in the GMO sector.

Measurement Uncertainty

Codex is currently developing guidelines on Measurement Uncertainty. These guidelines, as well as the accreditation requirements cited above, require laboratories to estimate the uncertainty of their quantitative measurements. This is particularly important and has consequences for measurements in the GMO sector where analytical controls may not be as effective as found in other areas of analysis in the food sector.

INFORMATION TO BE PROVIDED TO CODEX WHEN A METHOD FOR GMOs IS TO BE CONSIDERED FOR ENDORSEMENT BY CCMAS

The information that should be supplied to CCMAS when a method is to be considered for endorsement is given in Annex I. The annex lists both general considerations and specific requirements.

As GMO methodology becomes more developed the specific requirements will be converted to performance criteria to conform to the "criteria approach" already adopted by Codex.

DEFINITIONS

There are a number of Codex definitions applicable to GMO analysis. Suggested definitions are given in Annex II.

METHOD DEVELOPMENT TO FORMAL VALIDATION

Applicability of the Method

This is a particularly important criterion in GMO analysis. In principle the method should be applicable to the matrix of concern within the Codex system. If this a specific product derived from GMO then there is merit in requiring those seeking endorsement to provide information on the method of analysis appropriate to the specific product and, ideally, the matrix in which it is likely to be used. In case of "general purpose" GMO methods, at least one extraction method applicable to a general matrix should be available.

It is required from an extraction method, independent of matrix to which it is to be applied, that it yields DNA of sufficient quantity, structural integrity and purity to allow a proper evaluation of the performance of the subsequent method steps (e.g. proper amplification of the real-time PCR step) to be undertaken. This can be tested by setting up dilution series of the template DNA and see that the ΔCT between the dilutions correspond to the dilution factor, e.g. if DNA is diluted 10X then the ΔCT should be approx. 3.32, if the DNA is diluted 4X, the ΔCT should be 2, etc. Deviations from this relationship may indicate that the extracted DNA contains PCR inhibitors, that the DNA solution is not homogenous or the DNA quantity so low that stokastic variation in copy numbers yield unreliable quantitative estimates.

Validation Process

Method validation is a process of establishing the performance characteristics and limitations of an analytical method and the identification of the influences, which may change these characteristics - and to what extent. The results of a validation process describe which analytes can be determined in what kind of matrices in the presence of which interference. The validation exercise results in precision and accuracy values of a certain analytical method under the examined conditions.

Formal validation of a method is the conclusion of a long process, which includes the following main steps:

- ***Method development and optimisation.*** Prior to any pre-validation, the method should be fully optimised so that an inter-laboratory transfer is possible. The protocol should be finalized so that no major changes are needed between the pre-validation and validation.
- ***Pre-validation of the method***. Pre-validation should ensure that a method performs in a manner, which allows a successful conclusion of the validation study, i.e. it should provide evidence about the compliance with the regulations. Pre-validation should preferably be carried out by involving 2 – 4 laboratories.
- ***Full validation of the method.*** Full validation requires considerable resources and should be conducted only on methods which have received adequate prior testing.

A collaborative trial is expensive to undertake and usually follows only after the method has shown acceptable performance in a pre-validation study or Single-Laboratory study.

Modular Approach to Method Validation

The "method" refers to all the experimental procedures needed to estimate the measurand in a particular matrix. For a particular material this may include the methods for DNA extraction and the final quantification in a PCR system. In such a case, the whole chain from extraction up to the PCR-method (or equivalent) constitutes a method, but the different method parts can be considered separately (i.e. modular validation). In practice this is difficult to achieve.

The theoretical advantage of a modular approach to method validation is that each section of a method or protocol can be validated separately, and once validated, can be combined with other sections in a flexible manner.

However, there are several disadvantages to a modular approach to method validation, particularly when GMO analysis is being considered. It has been found that the variability of GMO analysis is very significant, and this then reduces the effectiveness in comparing different approaches to the same module in a method. But most critically, a modular approach to method validation assumes that the modules in a method which form the whole are independent of each other. This is frequently not the case and where "official control" work is to be undertaken, a modular approach should not be taken.

METHOD ACCEPTANCE CRITERIA

In order to evaluate a method prior to full validation, information concerning both the method and the method testing is required. Information on this is given in Annex I.

The method will be evaluated based on the information provided to Codex. The evaluation should verify that the principle preconditions for using the method for Codex purposes are fulfilled. This section describes the method acceptance criteria, which have to be fulfilled by the method in order to conduct further a pre-validation and full collaborative trial.

Principle Conditions

The provision of the detection method is aimed to serve mainly the requirements for the monitoring and labelling of GMOs, as set out in the specific regulations. To serve these purposes, the method should detect and quantify the specific GM event in the GM product. Currently, the detection method typically consists of a PCR-based method and includes:

- a protocol describing an extraction method which is applicable to a relevant matrix;
- a description of the primer sequences which uniquely identify the GM event in the GM product[1];
- a description of the primer sequences which amplify an endogenous gene sequence applicable to the specific host species;
- a protocol describing the conditions under which PCR can be used to detect the GM product;
- appropriate control samples.

The method provider should demonstrate that the method fulfils the principle method requirements:

(1) The method should be event-specific and allow for unequivocal detection/identification/ quantification of a specific event.

Currently, the best choice concerning event-specificity of a method, should PCR be the chosen technique, is to target an event-specific genomic region using a set of oligonucleotides (primers) that trigger the amplification of such a region. Among various types of event-specific genomic regions, the one relative to the junction between the transgenic insert and the host genomic DNA will probably be the location of choice. However, when a unique DNA sequence can be found within the transgenic insert, such a sequence can also be targeted by appropriate oligonucleotides and multiplied through a PCR.

(2) The method should be applicable to the material specified in its scope, and to appropriate quality control and reference materials.

COLLABORATIVE TRIAL REQUIREMENTS

[1] *Note: the fact that most event-specific sequences are not publicly disclosed should be discussed by CCMAS.*

General Information

The purpose of a collaborative trial is to fully validate the data provided by previous testing in a pre-validation or a single laboratory exercise and to determine methodological precision in terms of repeatability and reproducibility.

The values of any performance parameters reported from validation studies must be interpreted and compared with care. The exact values and their interpretation may depend – besides the performance of the method - on the extent of the method (e.g. a real-time quantitative PCR only versus a method chain ranging from extraction to the real-time PCR quantification), experimental design applied, exact calculation forms used to determine the parameters and the approach used to detect and analyse outliers. In order to have meaningful "minimum performance requirements" the above factors must be treated appropriately and in a standardized manner.

For Codex purposes the ISO/AOAC/IUPAC harmonized protocol has been adopted.

Minimum Performance Requirements

In a collaborative trial, the method performance should comply with the relevant parts of the method acceptance criteria and fulfil the method performance requirements specifically set below for the collaborative trial. Thus, the collaborative trial confirms the results obtained during the previous method evaluation phases and provides additional information about the method performance in a multi-laboratory setting. In particular, the compliance with the criteria for sensitivity and repeatability standard deviation should be re-confirmed.

In addition to the method acceptance criteria, at least the method performance requirements listed in Annex I should be evaluated from the experimental data of a collaborative trial. First, the definition and thereafter the requirements are described.

The endorsed methods and their associated validation data will be revised on a regular basis as the scientific knowledge and experience gained in Single-Laboratory validation and collaborative trials evolve. These Guidelines will also be complemented with practical information about the operational steps of the validation process.

Collaborative Trial Test Materials

In principle, the method should be applicable to and tested on the matrix of concern (i.e. on which any specification has been made).

In other fora recommendations have been made that in case of "general purpose" GMO procedures (in contrast to consideration of a specific product derived from GMO) that the validation of the detection module is carried out using genomic DNA as the analyte (for a PCR-based method). This allows the detection step to be combined with various extraction methods applicable to different matrices. However, real materials/matrix typical of a type/group of matrices are preferred and such a modular approach is inappropriate when considering Codex specifications.

VALIDATION OF PCR METHODS

Specific information on the validation of quantitative and qualitative PCR methods is given in Annexes III and IV respectively.

MEASUREMENT UNCERTAINTY

DNA-based analysis is commonly performed using Polymerase Chain Reaction (PCR). This technique amplifies a specific (short) segment of DNA to the extent that its quantity can be measured instrumentally (e.g., using fluorometric means). As DNA is a molecule that is easily degraded during food processing operations (e.g., due to heat, enzymes and mechanical shearing), we urge that this be considered in the performance criteria assessment of this technique. This is relevant as in most foods raw ingredients are not present, but are in a processed form, which has an effect on proteins and/or DNA present in food. Furthermore, these protein(s) and/or DNA may be degraded, or its total amount may be decreased due to processing. As a result, any current detection method (DNA- or protein-based) is affected.

It is often the case that the results of a determination are expressed in terms of percent of a sample that contains a particular biotechnology-derived sequence. In a quantitative test, this measurement actually involves two PCR-based determinations – that of the primary analyte (e.g. an inserted gene sequence) and that of the endogenous, or comparator sequence (e.g. an endogenous maize gene). Each of these

determinations has its own uncertainties, and the two are likely to have different measurement characteristics. In most applications, the primary analyte will be present at low concentrations, and the comparator will be present at concentrations 10 to 1000 times higher. It is thus important that both measurements are properly validated. In cases where the measurement is expressed directly as a percentage (as in the use of ΔCT), these factors must be considered when validating the method.

The consequence is that the analysis of DNA, especially in processed foods, aims at detecting a very small amount of analyte. Although the result of a PCR analysis is often expressed in % as the relative amount of DNA specific for foods derived from modern biotechnology relative to the total amount of DNA for a specific species, the actual amount of DNA specific for foods derived from modern biotechnology is often in the nanogram/gram range or lower. Analysis of those low amounts of analyte is accompanied by a considerable measurement uncertainty; this needs to be appreciated by the users of analytical results.

UNITS OF MEASUREMENT

Various countries have thresholds established for labelling of biotech- derived food and feed products. These thresholds are explicitly or implicitly expressed as weight by relative percentage. However, none of the current detection methods (DNA – or protein-based) are able to measure this directly. Although there is a correlation between weight-% and the amount of DNA or protein, respectively, the very nature of this relationship is influenced by a number of biological factors and thus remains highly variable. This continues to cause considerable misunderstanding and requires significant technical guidance.

This issue needs to be appropriately addressed and performance and data reporting criteria established for these methods.

GUIDANCE ON LABORATORY SET-UP AND OPERATION

DNA-based methods for the analysis of foods derived from modern biotechnology apply techniques that are not considered as commonly available methods, as they require specific apparatus and handling techniques that differ from most chemical-analytical methods. It is therefore necessary to provide information and instructions on the essential differences in laboratory set-up and handling techniques. Examples are available[2].

REFERENCE MATERIALS

There are a number of matrices that can be used to develop reference materials or working standards for methods of detection of GM products. Each has its own advantages and drawbacks for particular purposes.

At the present time there is insufficient availability of "pure" reference materials.

Codex may consider requiring the availability of suitable reference materials as part of the method endorsement procedure.

In the short-term the availability of quality control materials from proficiency testing schemes or from the use of Plasmid or amplicon DNA should be investigated.

SAMPLING

In the area of GMO analysis it may be anticipated that sampling error can be expected to contribute significantly – if not dominate - the overall uncertainty of an analytical result, particularly when considering raw commodities. The combination of sampling and analytical uncertainties is now being addressed by a number of International Organisations, most notably EURACHEM which has set up a new Working Group dealing with uncertainty of sampling. Much work has been carried out on sampling generally by CCMAS[3] and of bulk sampling for GMOs by the EU JRC[4] and GIPSA.

REFERENCES

To be completed.

[2] See draft ISO-standard (ISO/DIS 24276) or the corresponding French standard (AFNOR XP V03-020-2, tabled as room document CRD 5 in its previous version AFNOR XP V03-020-1 by the French Delegation at the 24th Session of CCMAS)

[3] Codex General Guidelines on Sampling.

[4] FP5 KeSTE project.

ANNEX I: INFORMATION TO BE PROVIDED TO CODEX WHEN A METHOD IS TO BE CONSIDERED FOR ENDORSEMENT BY CCMAS

In order to aid the endorsement of a proposed method of analysis in the GMO sector by Codex, and in particular CCMAS, the following should be provided:

Description of the method

A complete and detailed description of all the components of the method, and address also the following:

Purpose and relevance of the method: The objective of the method and the relevance of the method with respect to relevant legislative requirements should be indicated. In particular, the proposer should indicate that the principle conditions for the method are fulfilled.

Scientific basis: An overview of the principles of how the method, such as DNA molecular biology based (e.g. for real-time PCR) information should be provided. References to relevant scientific publications are useful.

The prediction model adopted to interpret results and to make inferences must be described in full details.

Specification of the prediction model/mathematical model needed for the method: If the derivation of the results relies upon a mathematical relationship this must be outlined and recorded (e.g., a regression line or calibration curve obtained by other means). Instructions for the correct application of the model should be provided. These may include, depending on the method, a recommended number and range of levels to be analysed, minimum number of replicates to be included or the means to evaluate the goodness-of-fit.

Outline of the experimental design, including the details about the number of runs, samples, replicates etc.

Information about the method optimisation

Primer pairs tested: Sufficient justification should be given of how and why the proposed primer pair has been selected, also for the reference gene (should this be part of the method).

Specificity testing: Empirical results from testing the method with non-target transgenic events and non-transgenic plants should be provided. This testing should include closely related events and cases were the limits of the sensitivity are truly tested.

Stability testing: Empirical results from testing the method with different varieties should be provided in order to demonstrate, for instance, the stability of the copy number of the reference gene.

Sensitivity testing: Empirical results from testing the method at different concentrations in order to test the sensitivity of the method. Limits of detection must be defined using samples comprising of single crops only, e.g. "the LOD for Roundup Ready® soy is 0.1 % of total soy *if the product is comprised of 100 % soy*". For food products made up of multiple ingredients, the actual sensitivity will be reduced, as total extracted DNA will be derived from more than one ingredient so that the starting amount of the actual measurand will be decreased. This dilution effect will depend on how much of the target ingredient (e.g. soy) is in the food product and the total quantity of DNA derived from the other ingredients. Some ingredients will contribute much DNA, such as wheat or maize flour and eggs, while other ingredients will not contribute any DNA, such as sugar, water or highly processed oils.

Robustness testing: Empirical results from testing the method against small but deliberate variations in method parameters.

Practical application of the method

Applicability: Indication of the matrix (e.g., processed food, raw materials, etc.), the type of samples (e.g., seeds, flour, pizza, cookies, etc.) and the range to which the method can be applied. Relevant limitations of the method should also be addressed (e.g. inference by other analytes or inapplicability to certain situations). Limitations may also include possible restrictions due to the costs, equipment or specific and non-specific risks implied for either the operator and/or the environment.

Operational characteristics and practicability of the method: The required equipment for the application of the method should be clearly mentioned, with regards to the analysis *per se* and the sample preparation. An indication of costs, timing, practical difficulties, and of any other factor that could be of importance for the operators should be also indicated.

Operator skills requirements: A description of the practical skills necessary to properly apply the proposed method should be provided.

Analytical Controls

The proper use of controls when applying the method should be indicated. Controls should be clearly specified and their interpretation recorded. These may include positive and negative controls, their detailed contents, the extent into which they should be used and the interpretation of the obtained values.

Positive and negative controls used

Control samples, plasmids and alike used

Reference materials used.

Method Validation/Performance

See the Codex "Check-list" (i.e. accuracy, applicability (matrix, concentration range and preference given to 'general' methods), detection limit, determination limit, precision, recovery, selectivity, sensitivity and linearity)

and in particular the following information should be supplied:

- amplicon length
- whether the method is instrument specific
- whether there are differences between qualitative and quantitative PCR-based detection methods
- whether there are differences between PCR-based and immunological methods concerning validation criteria
- whether single- or multi-plex PCR amplifications are undertaken
- maximum number of PCR cycles

The information provided should demonstrate the robustness of the method for inter-laboratory transferability. This means that the method should have been tested by at least one external laboratory besides the laboratory which has developed the method. This is an important pre-condition for the success of the validation of the method.

ANNEX II: CODEX DEFINITIONS APPLICABLE TO GMO ANALYSIS

This Annex is concerned with the definitions needed in GMO analysis. ***(Note: a number of definitions have been grouped together in one heading; these may be contradictory and this needs to be resolved. The Codex definition given in the Procedural Manual should be used and amplified as necessary. Codex definitions have not been reproduced here if they need no further qualification for GMO analysis)***.

Accuracy

The closeness of agreement between a reported result and the accepted reference value[5].

The accuracy should be within ± 25% of the accepted reference value over the whole dynamic range.

Applicability

The analytes, matrices and concentrations for which a method of analysis may be used[6].

The analytes, matrices, and concentrations should be appropriate for the control purposes for which the method has been proposed. The description may also include warnings to known inferences by other analytes, or inapplicability to certain matrices and situations.

It is not feasible to provide reference materials for every one of the many food matrices that are available, so that the use of a representative matrix reference will usually be necessary. The use of the method in a new matrix will need to be validated at a minimum via in-house validation, usually by spike and recovery experiments, and the reference material used should be described on the report to the customer.

Dynamic Range - Range Of Quantification

The interval of concentration within which the analytical procedure has been demonstrated by collaborative trial to have a suitable level of precision and accuracy.

The dynamic range of the method should cover at least 10% and 200% of the target concentration. Target concentration should be understood here as the threshold relevant for a certain regulation.

Example: 0.1% and 2.0% for a 1% GMO concentration or 50 and 1000 genome copies if the target is 500 copies.

There is a general scientific discussion still going on about the interpretation of the percentage values (e.g. dynamic range from 10 to 200% of the target value). Although the experts agreed that – at least for PCR – copy number is desired over weight/weight percentage, it was recognised that so far there is no reliable weight/copy number relationship because of inter-variety fluctuation of the 1C value and because of uncertainty in the correlation of weight of ingredient to weight of DNA. For the time being, both the w/w and copy number/copy number calculations are acceptable.

The unique characteristics of quantitative PCR impose particular restrictions on the low end of the dynamic range of a quantitative PCR. This is due to the difficulty in determining LOD and LOQ values due to the non-normal distribution of variances in the values in this range. Thus it may not be appropriate to require a range extending to 10% of the measured value. The suggestion of a dynamic range that ranges from 10 to 200% can be problematic. For example, capability to analyse a foodstuff composed of more than 50% (w/w) of a biotechnology-derived material (as might be the case for a nutraceutical) would require a dynamic range exceeding 100% (w/w). This is clearly not possible.

Limit of Detection (LOD)

(A)

Limit of detection is the lowest concentration or content of the analytes that can be detected reliably, but not necessarily quantified, as demonstrated by collaborative trial or single-laboratory validation[7].

Criteria: LOQ < 1/20th of the value of specification with an RSDr $\leq$ 33. The value of specification should be understood here as the threshold relevant for a certain regulation.

Note: limits of detection must be defined using samples comprising of single crops only, e.g. "the LOD for Roundup Ready® soy is 0.1 % of total soy *if the product is comprised of 100 % soy*". For food products

[5] Definition adopted from ISO 3534-1.

[6] Slightly modified from the definition provided in Codex CX/MAS 02/4: Proposed draft guidelines for evaluating acceptable methods of analysis. Version November 2002.

[7] Slightly modified from prEN ISO 24276:2002 (E).

made up of multiple ingredients, the actual sensitivity will be reduced, as total extracted DNA will be derived from more than one ingredient so that the starting amount of the actual measurand will be decreased. This dilution effect will depend on how much of the target ingredient (e.g. soy) is in the food product and the total quantity of DNA derived from the other ingredients. Some ingredients will contribute much DNA, such as wheat or maize flour and eggs, while other ingredients will not contribute any DNA, such as sugar, water or highly processed oils.

(B)

It is worth noting that a determination of an LOD or LOQ is not necessarily needed to establish the validity of a method for a given application. For example, it does not add much value if an LOD is determined to be 1ng/kg, while the scope of the method validation extends only for concentrations ranging in g/kg. In this and similar cases the reliability of the method will be proven by the other parameters and no efforts are included in the method validation to assess the LOD. However, the LOQ shall always be established and included in the validation study.

If the LOD is required, it is common practice to assume that it is the signal strength of a blank increased by three times the standard deviation of the blank. However, this method gives at best an estimate, relies on normal Gaussian distribution of the blank measurements around zero, and may give a lower value than the actual LOD. Its use is not valid in methods such as Quantitative PCR, in which the distribution of measurement values for blanks is typically truncated at zero and is thus not normally distributed. Thus the LOD need to be experimentally determined unless the targeted concentrations are well above the LOD and the LOD therefore becomes irrelevant. For quantitative methods the LOD is the amount of analyte at which the analytical method detects the presence of the analyte at least 95% of the time ($\leq$5% false negative results). This, and the false positive rate, are the only parameters required for a qualitative method other than specificity.

Limit of Quantification (LOQ)

(A)

The limit of quantification of an analytical procedure is the lowest amount or concentration of analyte in a sample, which can be quantitatively determined with an acceptable level of precision and accuracy as demonstrated by satisfactory collaborative trial or single-laboratory[8] validation[9].

Criteria: LOQ < 1/10th of the value of specification with an $RSD_r \leq 25\%$. The value of specification should be understood here as the threshold relevant for a certain regulation

Example: For a 1 % nominal value LOQmin = 0.1 % or for 500 copies LOQmin = 50 copies.

(B)

For a quantitative method, it is important to know whether the LOQ for a particular matrix is close to the values to be measured. Using the traditional approach, the LOQ can be expressed as the signal strength of a blank equal to the LOD increased by 6-10 times the standard deviation of the blank, unless it is known from other sources that the measured values range so high above the LOQ that its knowledge becomes irrelevant. However, this method to determine the LOQ leads only to an estimate of the true LOQ that may be an artificially high or low approximation.

In practice, two procedures have been employed to determine the LOQ. The first approach is to assay a number of negative samples that have been supplemented (spiked) with known amounts of analyte. The LOQ is then the level at which the variability of the result and percent recovery of the analyte meet certain preset criteria. For small molecules, these criteria have typically been a CV of 20% and 70-110% recovery [EPA ref.]. DNA recovery, however, may be difficult from some matrices, e.g. starches or ketchup, and lower recovery efficiencies may have to be accepted. When recovery efficiencies are low, this must be stated in the validation data and in the analytical report. A more complete approach is to test the method using a number of samples that contain known amounts of the GM material. This is more complicated as it requires access to significant quantities of reference materials that contain a known range of concentrations of the GM event of interest. Procedures for assessing LOD and LOQ during the validation of qualitative and quantitative PCR methods are discussed in Annex II.

[8] E.g. Thompson et al. 2002. IUPAC Technical Report: Harmonised guidelines fro single-laboratory validation of methods of analysis. Pure Appl. Chem. 74(5): 835-855.

[9] Slightly modified from prEN ISO 24276:2002 (E).

Validation of methods consists of two phases. The first is an in-house validation of all of the parameters above except reproducibility. The second is a collaborative trial, the main outcome of which is a measure of the repeatability and reproducibility together with detailed information on the transferability of methods between laboratories. It is strongly recommended that a small-scale collaborative trial be performed to test the general ruggedness of a particular method before the expense of organizing a large-scale trial is incurred. In case any improvement of the method or the method description are needed, only limited expenses are incurred through the pre-trial, while a failure of a full interlaboratory method validation due to a ambiguous method description is a very costly failure. Additionally, it may be pointed out that the implementation of an already validated method in a laboratory needs to include necessary experiments to confirm that the implemented method performs as well under local conditions as it did in the interlaboratory method validation. It is important to note that a method should be validated using the conditions under which it will be performed. Thus it should not be tested using an unreasonable number of amplification cycles. Most PCR reactions can be expected to result in analytical artefacts if operated with too many cycles and/or under non-optimal conditions.

Practicability

The ease of operations, in terms of sample throughput and costs, to achieve the required performance criteria and thereby meet the specified purpose[10].

Generally, the method should be practical for its intended purposes.

Repeatability standard deviation (RSD_r)

The standard deviation of test results obtained under repeatability conditions. Repeatability conditions are conditions where test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time.[11]

Repeatability standard deviation describes the intra-laboratory variation. The relative repeatability standard deviation should be below 30% over the whole dynamic range of the method.

Reproducibility standard deviation (RSD_R)

The standard deviation of test results obtained under reproducibility conditions. Reproducibility conditions are conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.[12]

Reproducibility standard deviation describes the inter-laboratory variation. The relative reproducibility standard deviation should be below 33% at the target concentration and over the majority of the dynamic range. $RSD_R < 50$ % at the limit of quantification/lower end of the dynamic range.

Robustness

Robustness refers to variations in the method as performed in different laboratories by different technicians. The language used here is derived from "Ruggedness" which is the equivalent in the harmonized guidelines. Ruggedness should be demonstrated by the validation of the method in 8-12 laboratories as defined in the harmonized guidelines. It is preferable from a CODEX point of view, that these laboratories be distributed across several continent/trading blocks.

The robustness of an analytical method is a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage[13].

The adequacy of the robustness testing needs to be analysed on method-by-method basis. For instance, for a real-time PCR method, following factors should be taken into account: at least two different thermal cycler models, DNA polymerase, uracil-n-glycosylase, magnesium chloride concentration, primer forward and reverse concentration, probe concentration, temperature profile, time profile, dNTP including dUTP concentrations.

[10] Adopted from prEN ISO 24276:2002 (E).
[11] Definitions adopted from ISO 3534-1.
[12] Definitions adopted from ISO 3534-1
[13] Definition adopted from ICH Topic Q 2 A "Validation of analytical methods: definitions and terminology." The European Agency for the evaluation of medicinal products. CPMP/ICH/381/95. Version November 1994. http://www.emea.eu.int/pdfs/human/ich/038195en.pdf

Sensitivity

The sensitivity of a method is a measure of the magnitude of the response caused by a certain amount of analyte.

The method should be sensitive enough in order to be able to detect/quantify with respect to the thresholds as provided in the relevant legislation.

Since sensitivity is method- and purpose-dependent it should be specified in the protocol. A reasonable goal for sensitivity is that required to meet levels specified in contracts, with a reasonable certainty that the level does not exceed the required limit. Typical sensitivities are in the range of 0.1% biotechnology-derived material by weight.

Sensitivity as a term is used in two different ways - LOD and the slope of a curve. The use of "detection limit", or "limit of detection" is the preferred term to use as a measure of the ability of a method to detect a small amount of analyte. See also previous comments regarding sensitivity in this document.

Specificity

(A)

Property of a method to respond exclusively to the characteristic or analyte of interest.

The specificity should be demonstrated by showing experimental results from testing the method with non-target transgenic events and non-transgenic plants. This testing should include closely related events and cases where the limits of the detection are truly tested. As the method should be event-specific it should only be functional with the GMO or GM based product considered and ought not to be functional if applied to other events already authorised. In addition, if a reference gene system is a part of the method this should not recognize any gene corresponding to even phylogenetically related species, and should give similar CT-values when amplifying equal amounts of DNA from different cultivars of the same species.

The adequacy of the testing needs to be analysed on a method-by-method basis. It will be necessary to obtain information about the specificity testing in case of stacked genes at some stage.

(B)

Specificity is the starting point for a method and needs to be considered during primer design. Primers should be checked against the known sequence of the event insert and pertinent databases for possible matches. Specificity must also be demonstrated experimentally. The following suggests a reasonable approach and the experiments should be performed during pre-validation of an assay.

For event-specific assays:

- Analyse at least a total of ten non-target transgenic events and any non-transgenic plants that may commonly be found as contaminants in the commodity.
- Test on sample from each source (total of at least 10 DNA samples).
- Analyse two replicates for each DNA sample.

Test results shall clearly indicate that no significant instrument reading is observed.

For assays on plant endogenous genes:

- Analyse at least a total of ten different plant samples comprising different varieties of the same plant species as well as other plants species important for food production (such as wheat, rice, corn, potato, and soybean) and that may commonly be found as contaminants in the commodity.
- Test one sample from each source (total of at least 10 DNA samples).
- Analyse two replicates for each DNA sample.

Test results shall clearly indicate that no significant instrument reading is observed.

Trueness

The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value[14].

The measure of trueness is usually expressed in terms of bias. Trueness has also been referred to as "accuracy of the mean".

The trueness should be within ± 25% of the accepted reference value over the whole dynamic range. This refers to the PCR-step provided that a modular validation approach has been applied.

[14] Adopted from ISO 3534.

ANNEX III: VALIDATION OF A QUANTITATIVE PCR METHOD

Introduction

A quantitative PCR assay should be validated for the intended use or application. A harmonized ISO/IUPAC/AOAC protocol was developed for chemical analytical methods. This defines the procedures necessary to validate a method (Horwitz W; Protocol for the design, conduct and interpretation of method-performance studies. Pure and Applied Chemistry, 67, 331 (1995)). It is important to emphasize that all the principles and rules of the harmonized protocol are applicable to quantitative PCR methods.

A number of the parameters involved in validation of the performance of a quantitative PCR assay will be discussed in detail. These are scope, LOD and LOQ, accuracy, precision, sensitivity and ruggedness (robustness). Other important factors are acceptance criteria and interpretation of results, and the issue of the units in which results are expressed.

It is important to note that a quantitative PCR assay typically consists of two assays, one determines the amount of DNA specific for the transgenic product, while the other is specific for the amount of plant specific DNA. Each of these assays has to be considered separately, as these assays can be considered as independent analytical procedures. Thus, all parameters listed below, including specificity and sensitivity, have to be assessed individually for each of the assays involved. A method validation for the whole assay cannot be appropriately performed unless both assays are validated individually.

Scope (Range of Quantification)

The scope of the methods defines the concentration range over which the analyte will be determined. Typically the range for a GM product will range from a tenth of a percent up to a few percent and for the endogenous control the range will be close to 100%, unless the testing of complex mixtures is envisioned. This desired concentration range defines the standard curves and a sufficient number of standards must be used to adequately define the relationship between concentration and response. The relationship between response and concentration should be demonstrated to be continuous, reproducible and should be linear after suitable transformation.

The range of a quantitative method is typically designed to be in the range 0.1% - 100% (DNA %w/w). However, it is common to validate a method for a range of concentrations that is relevant to the scope of the application. If a method is validated for a given range of values, the range may not be extended without validation. For certain applications (e.g., seed or grain analysis) the use of genomic DNA for the preparation of the standard curve (see discussion on the use of plasmid DNA below) may be considered. While it is easy to establish a nominal 100% standard (limited only by the purity of the materials used) it is difficult to reliably produce standard solutions below 0.1%. This is due to the uncertainties involved in measuring small volumes and the error propagation if serial dilution steps are applied. Additionally, the number of target sites (DNA sequence to be amplified) becomes so small that stochastic errors will begin to dominate and no reliable analysis is possible. (Huebner P, Waiblinger H U, Pietsch K, Bordmann P (2001) Validation of PCR methods for quantitation of genetically modified plants in food. Journal of AOAC International 84(6) 1855-1864; Kay S, Van den Eede G, The limits of GMO detection, Nature Biotech. 19(5) 504 (2001)). If genomic DNA is chosen to be used as calibrator, it is important that this calibrator needs to be traced back (in its metrological meaning) to a reference of highest metrological order, e.g. a certified reference material. The range will be established by confirming that the PCR procedure provides an acceptable degree of linearity and accuracy when applied to samples containing amounts of analyte within or at the extremes of the specified range of the procedure.

LOD and LOQ

If the validation of the quantitative PCR assay shows that the assay can measure transgenic plant DNA at 0.1% with acceptable trueness and precision, then it is often not necessary to determine the LOD and LOQ, as the method is only being applied above the range where these are relevant. However, if the method is being used at concentrations close to the limit of detection and limit of quantification (typically 0.01-0.05%), then the assessment of the LOD and LOQ will become part of the validation procedure.

Accuracy and Precision

As for any method, the accuracy of a method should be compared to known values derived from reference materials. This may be a challenge in this field due to the limited availability of such materials. However, the accuracy must be compared to the best available reference material. Precision will be determined in the usual way from single laboratory (repeatability) and multi-laboratory (reproducibility) studies.

Sensitivity

A linear relationship of the CT as a function of the logarithm of the concentration of the target of the individual target should be obtained across the range of the method. The correlation coefficient, y-intercept, slope of the regression line and % of residual should be reported. The % of residual for each of the calibrators should preferably be $\leq$=30%.

In order to obtain a standard curve for event specific quantitative assays, standard DNA mixtures can be prepared by combining purified genomic DNA from transgenic and non-transgenic plants material such as seed or leaves. The content of transgenic plant DNA in the mixtures might be 100, 50, 10, 5, 1, 0.5, 0.1, and 0% or as appropriate for a smaller concentration range. Three replicates must be analysed for each point on the standard curve.

For quantitative assays on plant endogenous genes, standard DNA mixtures can be prepared by combining purified genomic DNA from the target plant species and that of a non-target plant species. For example, for validation of a maize ADH1 quantitative assay, the target plant species is maize and the non-target plant species could be soybean or another species. The content of DNA of the target plant species in the mixtures is typically 100, 50, 25, 10, 5, 1 and 0% or as appropriate. Three replicates must be analysed for each point on the standard curve.

In cases where the ΔCT-method is employed, it will be the responsibility of the analyst to ensure that the overall amount of DNA is well within the range for which the assay was validated.

Ruggedness (Robustness)

The evaluation of ruggedness (robustness) demonstrates the reliability of a method with respect to inadvertent variation in assay parameters. Variations that may be included include reaction volumes (e.g., 25 vs. 30µl), annealing temperature (e.g., plus and minus 1^{o}C) and/or other relevant variations. The experiments need to be performed at least in triplicates and the recovery needs to be calculated. The response of an assay with respect to these small changes should not deviate more than ±30% from the response obtained under the original conditions.

Analytical Control Acceptance Criteria and Interpretation of Results

A validated method also includes criteria on which the validity of an observed measurement result can be accepted as valid. It is important to follow these criteria and to observe the rules for data interpretation. In the case that it may be desired to deviate from said criteria and rules a new method validation study would be needed in order to demonstrate the validity of the new rules and procedures.

At a minimum, the following acceptance criteria are common to all quantitative PCR methods and applicable to each PCR run:

- The result of the positive DNA target control, with, for example 1% transgenic DNA, the mean of the replicates deviates less than 3 standard deviations from the assigned value. A target DNA control is defined as reference DNA or DNA extracted from a certified reference material or known positive sample representative of the sequence or organism under study. The control is intended to demonstrate what the result of analyses of test samples containing the target sequence will be.
- The amplification reagent control is $\leq$ LOD. The amplification control is defined as control containing all the reagents, except extracted test sample template DNA. Instead of the template DNA, a corresponding volume of nucleic acid free water is added to the reaction.
- The % of residual for each of the standards should be <30%

To accept the result of an unknown sample, the relative standard deviation of the sample replicates should be <30%.

ANNEX IV: VALIDATION OF A QUALITATIVE PCR METHOD

Introduction

A qualitative PCR that is used in a semi-quantitative approach must be validated in the same way as it is intended to be used – that is the sensitivity of the method must be shown to be such that it can reliably detect one positive particle (seed) in a pool, and does not give rise to a significant number of false positives. A concept of using false-positive and false-negative rates to describe the accuracy and precision of a qualitative assay has been developed for microbial assays (AOAC® Official Methods℠ Program Manual, Appendix X p14f, May 2002, AOAC International; http: www.aoac.org/vmeth/omamanual/htm). This concept can be applied to qualitative PCR assays. A critical issue in the validation of this type of method is the availability of test materials that are known to be positive and negative. The provision of negative reference materials is particularly important and critical in the case of a qualitative method. Any impurities must be present only at levels so low that they become negligible.

By their very nature, qualitative test results refer to the identification above/below a limit. The measures of precision and accuracy are the frequencies of false negative and/or false positive results at the detection limit. False negative results indicate the absence of a given analyte when in fact the analyte is present in the sample, while false positive results indicate the presence of an analyte that is not present in the sample. Due to the inherent nature of the analytical technique, an increase in false negative results will be observed when the amount of analyte approaches the LOD of the method. Like the limit of detection for quantitative methods, the limit of detection for a qualitative method can be defined as the concentration at which a positive sample yields a positive result at least 95% of the time. This results in a rate of false negative results of 5% or less. During validation of a qualitative PCR assay, it is also important to determine the number of false positive results (a positive result obtained using a sample that is known to be negative). This is also expressed as a rate.

False Positive Rate

This is the probability that a known negative test sample has been classified as positive by the method. The false positive rate is the number of misclassified known negatives divided by the total number of negative test samples (misclassified positives plus the number of correctly classified known negatives) obtained with the method:

For convenience this rate can be expressed as percentage:

$$\%\ \text{false positive results} = \frac{\text{number of misclassified known negative samples}}{\text{total number of negative test results [incl. misclassified]}}$$

Note: different sectors use different definitions here.

False Negative Rate

This is the probability that a known positive test sample has been classified as negative by the method. The false negative rate is the number of misclassified known positives divided by the total number of positive test samples (misclassified positives plus the number of correctly classified known positives) obtained with the method.

For convenience this rate can be expressed as percentage:

$$\%\ \text{false negative results} = \frac{\text{number of misclassified known negative samples}}{\text{total number of positive test results [incl. misclassified]}}$$

Note: different sectors use different definitions here.

In order to demonstrate the false negative rate for qualitative assay, a series of samples (e.g. grain/seed pools) with a constant, known concentration of positive material in a pool of negative material (e.g., 1 positive kernel in 199 conventional corn kernels) have to be analysed and the results evaluated. It is important to note that the concept of confidence intervals and statistical uncertainty needs to be applied to the risk of false positive and/or false negative results as well. The desired level of confidence determines the size and number of pools that need to be tested. For example, 100 positive test results obtained from 100 independent measurements on truly positive samples lead to the conclusion that the level of false negative results is below

4.5% at a confidence level of 99% for the tested concentration of positive kernels (expressed as the number of positive kernels in a pool of negative kernels).

Ruggedness

As with any validated method, reasonable efforts must be made to demonstrate the ruggedness of the assay. This involves careful optimisation and investigation of the impact of small modifications that could occur for technical reasons.

Acceptance Criteria and Interpretation of Results

A validated method includes criteria on which the validity of an observed measurement result can be accepted as valid. It is important to follow these criteria and to observe the rules for data interpretation. It is therefore important to make sure that the result of the positive DNA target control, is positive. Similarly the amplification reagent control must be negative. In addition to these controls, it is desirable to carry out a parallel reaction on the same DNA sample using a primer set which detects an endogenous single copy sequence. This reaction is carried out on every DNA sample, and can either be in the same reaction (multiplexed) or as a separate reaction. In the case of multiplexed reactions, it is important that the endogenous reaction does not out compete the event specific reaction for reagents, as the endogenous sequence is likely to be present at up to 1000 fold the amount of the target sequence. The control reaction with the endogenous sequence gives an indication of the quality of the DNA as a template for the PCR reaction. Table 1 sets out the accept/reject criteria for the PCR reactions on a per lane basis, using the results of the PCR reaction with the endogenous sequence.

Table 1: Criteria for scoring Qualitative PCR analyses

PCR result (GM analyte)	PCR result (endogenous)	Scoring of test
+	+	+
-	+	-
+	-	+
-	-	Reject

A further complication is however introduced by the fact that qualitative PCR reactions are typically carried out in duplicate. Thus it can occur that the duplicates do not agree. it is common practice to repeat PCR reactions once on DNA samples that are rejected. A repeated indeterminate result is indicative that the analyte cannot be reliably detected. (Table 2), and that the assay is operating below the limit of detection as, by definition, a 95% or better detection rate would be achieved at the limit of detection. The sample is therefore scored negative. Similar criteria apply if more replicates are carried out on each DNA sample.

Table 2: Criteria for scoring duplicate qualitative PCR analyses

Lane 1	Lane 2	Scoring of test
+	+	Positive
-	+	Repeat/Indeterminate
+	-	Repeat/Indeterminate
-	-	Negative