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[3].

Applicability

The analytes, matrices and concentrations for which a method of analysis may be used[4].

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 Single Laboratory 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.

Limit of Detection (LOD)

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[5]. LOD is generally expressed as the amount of analyte at which the analytical method detects the presence of the analyte at least 95% of the time (≤5% false negative results).

Limit of Quantification (LOQ)

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[6] validation[7].

Practicability

The ease of operations, in terms of sample throughput and costs, to achieve the required performance criteria and thereby meet the specified purpose[8].

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.[9]

Reproducibility standard deviation (RSD_R)

---

[3] Definition adopted from ISO 3534-1.

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

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

[6] 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.

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

[8] Adopted from prEN ISO 24276:2002 (E).

[9] Definitions adopted from ISO 3534-1.

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.[10]

Recovery

Proportion of the amount of analyte, present in or added to the analytical portion of the test material, which is extracted and presented for measurement.

Ruggedness (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[11].

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.

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

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

Trueness

The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value[12].

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

[10] Definitions adopted from ISO 3534-1

[11] 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

[12] Adopted from ISO 3534.

ANNEX III: VALIDATION OF A QUANTITATIVE PCR METHOD

INTRODUCTION

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.

VALIDATION

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. These are given alphabetically, not necessarily in order of importance.

Accuracy

As for any method, the accuracy of a method should be compared to known values derived from reference materials, ideally the best characterised. Precision will be determined in the usual way from single laboratory (repeatability) and multi-laboratory (reproducibility) studies.

Recommendation: The accuracy should be within [± 35%] 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 must be stated.

Dynamic Range - 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[1, 2]. 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.

Recommendation: The dynamic range of the method should cover at least [20% and 5] times the target concentration, where practicable. 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 5 times 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)

If the validation of the quantitative PCR assay shows that the assay can measure transgenic plant DNA at (for example) 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.

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 (≤5% false negative results). This, and the false positive rate, are the only parameters required for a qualitative method other than specificity.

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 70-120% recovery. 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 quantitative PCR methods are discussed in Annexes III and IV..

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.

Recommendation: Limit of detection is to be < 10% of the value of specification. The value of specification should be understood here as the threshold relevant for a certain application.

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 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.

Limit of Quantification (LOQ)

See introduction above for limit of detection.

Recommendation: The limit of quantification is to be < 20% of the value of specification with an $RSD_r \leq$ [25%] or as close as is practicable. 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.

For a quantitative method, it is important to know whether the LOQ for a particular matrix is close to the values to be measured. Traditional methods of approximating the LOQ (zero value plus 6-10 standard deviations) rely on normal Gaussian distribution of the blank measurements around zero. This approach 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 LOQ needs to be experimentally determined.

Practicability

The practicability of the method must be demonstrated.

Repeatability standard deviation (RSD_r)

Recommendation: The relative repeatability standard deviation should be below 25% over the whole dynamic range of the method.

Reproducibility standard deviation (RSD_R)

Recommendation: The relative reproducibility standard deviation should be below 35% 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.

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 are reaction volumes (e.g., 25 vs. 30µl), annealing temperature (e.g., plus and minus 1°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.

The adequacy of the robustness testing needs to be analysed on method-by-method basis. For instance, for a real-time PCR method, the following factors should ideally be taken into account: 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.

Sensitivity

For a quantitative PCR method, 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 ≤=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. Alternatively the 5% CRM used and further diluted without target DNA.

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.

Recommendation: Typical sensitivities are in the range of 0.1% biotechnology-derived material by weight if the material is not highly processed.

Specificity

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.

Recommendation: 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

Recommendation: The trueness should be within ± [30%] 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.

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 ≤ 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 <[35]%.

REFERENCES FOR ANNEX III

1. 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;

2. Kay S, Van den Eede G, The limits of GMO detection, Nature Biotech. 19(5) 504 (2001)).

3. Residue Chemistry Test Guidelines OPPTS 860.1340 "Residue Analytical Method" United States Environmental Protection Agency, August 1996, (Mihaliak & Berberich, 1995).

ANNEX IV: VALIDATION OF A QUALITATIVE PCR METHOD

Introduction

A qualitative PCR 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[1]. 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

REFERENCES FOR ANNEX IV

1. AOAC® Official Methods℠ Program Manual, Appendix X p14f, May 2002, AOAC International; http: www.aoac.org/vmeth/omamanual/htm.

ANNEX V: VALIDATION OF A PROTEIN-BASED METHOD[1]

QUANTITATIVE TESTING

Quantitative immunoassays are used to determine levels of the protein analyte in specific parts of the plant (e.g. seed, leaf, root, stalk etc). Typical applications are given in Table 1. In order to perform a microplate ELISA for quantitative determination of a protein analyte in plant tissue, it is first necessary to obtain a representative sample of the plant material. The sample amount will influence the detection limit or sensitivity of the assay. The analyte is then extracted from the plant material by adding a solvent and blending, agitating, or applying sheering or sonic forces. Typical solvents used are water or buffered salt solutions. Sometimes detergents or surfactants are added. Some proteins require more rigorous procedures like homogenisation or boiling in solvents, detergents, salts etc.

After the capture antibody has been immobilized on the microplate well surface, a precise volume of the standard or sample extract solution is added to each well. The analyte in the test solution binds to the capture antibody. The enzyme-labelled second antibody is then added and also binds to the analyte, forming a sandwich. At this point, the well is washed to remove unbound analyte and antibodies, leaving only the antibody-analyte-antibody complex bound to the well surface. A colorimetric substrate is added which reacts with the enzyme label and produces a coloured product. The reaction is stopped after a set period of time and the colour absorbance at a given wavelength is measured on a photometer. The standard curve is generated by plotting the optical density (OD) on the *y*-axis (linear scale) against the concentration on the *x*-axis (log scale) which produces a sigmoidal dose response curve Figure 4.

To obtain an accurate and precise quantitative value, the OD for the sample solutions must fall on the linear portion of the standard curve. If the OD is too high, the sample solution must be diluted until the OD falls within the quantitative range of the assay. The concentration of the protein analyte in the original sample of plant material is calculated by correcting for any dilution factor that was introduced in preparing the sample for application to the microplate. The initial weight of the sample and the volume of extraction solvent, as well as any subsequent dilutions are used to calculate the dilution factor.

Various assay controls can be employed to demonstrate the performance of the assay. A blank sample such as an empty well or buffered solution can be run in the assay to determine any background response which can be subtracted from sample and standard responses if desired. A negative control sample (i.e. matrix extract solution known to contain no analyte) can be used to demonstrate whether a non-specific response or matrix effect is occurring in the assay. A positive control or matrix extract spiked with a known amount of the analyte can be run to demonstrate accuracy. Standards and samples can be run in replicate to demonstrate precision. Blanks, negative controls, positive controls, fortified sample extracts, standardized reference material extracts, and replicates are typically run on each microplate to control for plate-plate variation.

STANDARDIZED REFERENCE MATERIALS

The standardized reference material consists of the same matrix as the actual agricultural commodity to be tested. For example, if the matrix to be tested is soybean seed, the standardized reference material would be soybean seed containing a known proportion of transgenic seed. Alternatively, a pure sample or extract of the protein of interest may be used, providing the use of such protein reference materials has been validated against the matrix in question. Access to standardized reference materials is important during the development, validation, and use of immunoassays for analysis of introduced proteins in transgenic agricultural commodities. The best available reference material should be used in order to comply with regulations and testing needs.

In the case of commodities such as grain or seed, where the commodity consists of discrete units, it is fairly straightforward to make a reference sample with a known proportion of transgenic material. In other cases, generating reference samples for certain matrices and analytes can be difficult. Stability and uniformity are important considerations. For example, if the matrix to be tested consists of a mixture of materials, it would be difficult to combine transgenic and non-transgenic material in such a way as to achieve a homogeneous reference sample with a known proportion of transgenic material. The stability of these materials would need to be evaluated under storage and test conditions. In any case, it is useful to have non-transgenic and transgenic material available to use as negative and positive controls.

During assay development, the reference material is used to help select assay parameters which would minimize any interfering effects of the matrix (e.g. non-specific binding of sample components to the

antibodies). During validation and use of the assay, the reference materials can be extracted and analysed alongside the test samples so that the results can be directly compared.

VALIDATION OF A QUANTITATIVE PROTEIN-BASED METHOD

The principles of method validation described in appendices III and IV for PCR methods also apply to protein methods. For commercially available immunoassay kits, assay performance is generally validated by the manufacturer and is documented in the product user's guide.

Quantitative protein-based methods are better characterized as a class than PCR-based methods. Validation should be conducted according to the harmonized ISO/IUPAC/AOAC protocol was developed for chemical analytical methods. This defines the procedures necessary to validate a method[2].

Accuracy: Accuracy is demonstrated by measuring the recovery of analyte from fortified samples and is reported as the mean recovery at several levels across the quantitative range. Ideally, quantitative methods will have demonstrated recoveries between 70 and 120% and a coefficient of variation (CV) of less than 20% for measured recoveries at each fortification level (Mihaliak & Berberich, 1995).

Extraction efficiency: Extraction efficiency is a measure of how efficient a given extraction method is at separating the protein analyte from the matrix. It is expressed as percent analyte recovered from the sample. Since the introduced protein expressed is endogenous to the plant, it can be difficult to demonstrate efficiency of the extraction procedure. There may not be an alternate detection method against which to compare the immunoassay results. One approach to addressing extraction efficiency is to demonstrate the recovery of each type of introduced protein analyte from each type of food fraction by exhaustive extraction, i.e. repeatedly extracting the sample until no more of the protein is detected (Stave, 1999).

Precision: Intra-assay precision describes how much variation occurs within an assay. it can be evaluated by determining the variation (%CV) between replicates assayed at various concentrations on the standard curve and on the pooled variation (%CV) derived from absorbance values in standards from independent assays performed on different days. Interassay precision describes how much variation occurs between separate assays and can be measured by analysis of quality control samples on every microplate. The quality control samples required would consist of two pools of extracts, one extract from transgenic plant tissue and one from conventional plant tissue. These extracts would be stored frozen and a portion would be thawed and assayed on every microplate. Interassay precision could be evaluated over time and expressed as % CV (Rogan *et al*, 1999). The precision of protein-based quantitative methods is in general higher than PCR-based methods.

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

Sensitivity: The sensitivity of the assay could be defined as the amount of analyte that can be measured by an absorbance reading of two standard deviations above background absorbance (Rogan *et al*, 1992). The detection limit could be expressed as the lowest dilution of transgenic crop that could be detected when transgenic and non-transgenic crop are combined (Rogan *et al*, 1999).

Dynamic Range - Range Of Quantification

The scope of the methods defines the concentration range over which the analyte will be determined. In most cases the analytical range for a GM product will range from a tenth of a percent up to a few percent. 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.

Interpretation of the percentage values (e.g. dynamic range from 10% to 5 times the target value) can be difficult when using quantitative methods. Quantitative protein methods generally give an estimate of the concentration of the GM protein in the matrix, due to variations in the expression of the amount of protein in different tissues of plants, and within the same tissue at different locations. The use of qualitative protein-based methods is thus much more prevalent. In addition, care must be taken to employ a method which can detect the protein in the matrix. For example, it is believed that proteins undergo modification or degradation due to processing to a greater degree than DNA, and thus loss of signal due to processing effects must be considered.

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.

Limit of Detection (LOD)

LOD is defined in annex II. Proteins are present in GM foods at higher concentrations than the target DNA for PCR methods. Thus stochastic effects have less influence on the determination of the LOD than when using PCR.

It is common practice when estimating the LOD to assume that it is the signal strength of a blank increased by three times the standard deviation of the blank. This method gives at best an estimate, and relies on normal Gaussian distribution of the blank measurements around zero. This can generally be assumed for methods such as ELISA, but the LOD is best determined experimentally. Alternatively the LOD is commonly defined as a concentration equal to the lowest standard used in the assay, should a positive value be consistently obtained with that standard.

Limit of Quantification (LOQ)

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 70-120% recovery[3]. Protein recovery, however, may be difficult from some matrices, e.g. starches or oils, 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 quantitative PCR methods are also discussed in annexes III and IV.

Specificity

The specificity is the degree to which analogs or other molecules bind to the antibodies and should be characterized and described in the method. 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 protein-specific it should only be functional with the GMO or GM based products considered and ought not to be functional if applied to events which do not express the protein in question. Interferences: the potential for interferences from reagents and labware can be evaluated by assaying extracts from non-transgenic plant material.

Matrix effects: if the response of the method is affected by a substance in the final extract other than the specific protein analyte, the non-specific response is referred to as a matrix effect. One way to manage matrix effects is to demonstrate that the analytical method gives identical results with or without sample matrix present in the extract. In this approach, freedom from matrix effects would have to be demonstrated in all matrices for which the assay is to be used. Another approach (although less desirable) to managing matrix effects would be to prepare the standard solutions in extracts from non-transgenic matrix, i.e. matrix-matched standards. This would ensure that any matrix effects would be consistent between the standards and the samples.

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 are reaction volumes incubation temperature (e.g., plus and minus 5-10°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. Experiments which may be performed to establish ruggedness include repeated analysis of a sample or samples on several days and measurement of accuracy and precision in fortified samples using control material from several sources.

QUALITATIVE (THRESHOLD) TESTING

Lateral flow devices are useful tools for on-site or field threshold testing. This type of testing requires a quick, accurate and cost-effective approach. In order to ensure reliable results, the manufacturer of the lateral flow device must conduct a method validation and provide a description of the performance characteristics of the product in the package insert. If this has been completed there is generally no need for validation studies to be performed by users. Each lateral flow device is an individual stand-alone unit, capable of performing to the standards described in the product package insert.

In order to establish an on-site procedure for threshold testing, the threshold level must first be established. To establish that the lateral flow device is able to differentiate between samples containing transgenic protein above or below the threshold, both a negative reference and a threshold reference containing a known proportion of transgenic grain should be assayed concurrently. The negative reference is a sample of the test matrix known to contain none of the protein analyte and is assayed to demonstrate that the method can distinguish between zero and the threshold level. A sufficient number of these samples are run to ensure that assay sensitivity is adequate to determine whether the level in the test sample is greater or less than the threshold level. During routine testing of bulk commodity samples, the lateral flow devices would typically be used without running the concurrent negative and threshold reference samples.

VALIDATION OF A QUALITATIVE (THRESHOLD) PROTEIN-BASED METHOD

The same principles apply to qualitative protein-based testing as to qualitative PCR testing. These approaches, including calculation of false positive and false negative rates, can therefore be applied to protein-based methods. In general, due to the more reliable nature of protein-based lateral flow strip methods, they are not performed in duplicate on each sample. However, if threshold ELISA testing is performed, duplicate wells should be used.

The same types of control samples, and criteria for acceptance/rejection of the result can be used as for qualitative PCR methods. The LOD is expressed as the amount of analyte at which the analytical method detects the presence of the analyte at least 95% of the time (<5% false negative results). However, lateral flow strip tests are generally applied at test concentrations that are at least two fold (or more) above the LOD.

REFERENCES FOR ANNEX V

1. Guidelines for the Validation and Use of Immunoassays for Determination of Introduced Proteins in Biotechnology Enhanced Crops and Derived Food Ingredients. Lipton et al., Food and Agricultural Immunology, 2000, 12, 153-164.

2. Horwitz W; Protocol for the design, conduct and interpretation of method-performance studies. Pure and Applied Chemistry, 67, 331 (1995).

3. Residue Chemistry Test Guidelines OPPTS 860.1340 "Residue Analytical Method" United States Environmental Protection Agency, August 1996, (Mihaliak & Berberich, 1995).