

Proficiency testing for quality assurance of allergens methods

Linda Owen · John Gilbert

Received: 20 December 2008 / Revised: 24 February 2009 / Accepted: 3 March 2009

© British Crown Copyright 2009

Abstract Proficiency testing is an external quality control check, whereby the quality of an analytical result is checked against criteria that are set independently of the laboratory carrying out the analysis. Participants in a proficiency test are encouraged to use the method of their choice to determine the analyte in question. The collated results submitted by the participants are used to derive the best estimate of the 'true' level, or assigned value, of the analyte, as a consensus value of the whole data set. Generally, the data submitted will be normally distributed and from a single population, but if a data set is found to be multimodal, then the selection of one of the modes as the assigned value is possible where there is supporting data, typically methodology information. Unless there are independent grounds for preferring one mode over another, it is not possible to set an assigned value or calculate *z*-scores. However, the analysis of allergens has presented proficiency testers with a new challenge, since it has become apparent that quantitative results may be dependent on the brand of enzyme-linked immunosorbent assay kit used, the specific analyte targeted (e.g. total content or allergen protein content) and the limit of detection achievable. FAPAS® has run more than 40 proficiency tests for allergen analysis over the past 7 years, during which time methods have been developed and improved and the requirements for determination of food ingredient allergens has increased. Two case studies are presented which highlight some of the issues around the use of allergen measurement methods.

Keywords Allergens · ELISA kits · Proficiency testing · Hazelnut · Sesame

L. Owen (✉) · J. Gilbert
Central Science Laboratory,
Sand Hutton,
York YO41 1LZ, UK
e-mail: l.owen@csl.gov.uk

Introduction

The incidence of food allergy in the population has apparently been increasing over the past decade [1, 2] and has been the subject of both national policies and European Council directives to help consumers make informed decisions over their food choices. Directive 2000/13/EC (as amended) of the European Parliament and the Council of 20 March 2000 required labelling of a specified list of allergenic ingredients as set out in Annex IIIa of that Directive. A number of applications for permanent exemption from the labelling requirement were submitted to the European Food Safety Authority (EFSA), and on the basis of EFSA opinions and other available information, it was concluded that certain ingredients are not likely to cause adverse reactions in susceptible individuals. This resulted in Commission Directive 2007/68/EC of 27 November 2007, amending Annex IIIa and listing food ingredients or substances derived from these specified allergenic ingredients that are no longer considered to cause adverse reactions in susceptible individuals (Table 1). The EFSA Scientific Panel on Dietetic Products, Nutrition and Allergies has commented [3] that analysis of foods for traces of potential food allergens is desirable, although major problems exist with regard to analytical techniques. Proficiency testing has revealed that results can be specific to the method of analysis and that acceptability of results must depend on the fitness for purpose criteria applied to the reason for analysis.

What is proficiency testing?

The analysis of an external quality check sample as part of a laboratory's routine procedures provides objective stand-

Table 1 Allergenic ingredients (from Annex IIIa, Commission Directive 2007/68/EC)

Number	Ingredients and products thereof	Exceptions
1	Cereals containing gluten	Wheat-based glucose syrups including dextrose Wheat-based maltodextrins Glucose syrups based on barley Cereals used for making distillates or ethyl alcohol of agricultural origin for spirit drinks and other alcoholic beverages
2	Crustaceans	
3	Eggs	
4	Fish	Fish gelatine used as a carrier for vitamin or carotenoid preparations Fish gelatine or isinglass used as fining agent in beer and wine
5	Peanuts	
6	Soybeans	Fully refined soybean oil and fat Certain tocopherols Vegetable oils derived from phytosterols and phytosterol esters from soybean sources Plant stanol ester produced from soybean sources
7	Milk (including lactose)	Whey used for making distillates or ethyl alcohol of agricultural origin for spirit drinks and other alcoholic beverages Lactitol
8	Nuts, i.e. almonds, hazelnuts, walnuts, cashews, pecan nuts, Brazil nuts, pistachio nuts, macadamia nuts, Queensland nuts	Nuts used for making distillates or ethyl alcohol of agricultural origin for spirit drinks and other alcoholic beverages
9	Celery	
10	Mustard	
11	Sesame	
12	Sulphur dioxide and sulphites at concentrations >10 mg/kg or 10 mg/l as SO ₂	
13	Lupin	
14	Molluscs	

ards for individual laboratories to perform against and permits them to compare their analytical results with those from other laboratories. Such standards and comparisons can go beyond the actual chemical analysis. For example, the ability to report results in specified units and within a given time scale are important aspects of quality. Proficiency testing (PT) is, therefore, an external quality control procedure. The quality of an analytical result is assessed against criteria that are set independently of the laboratory carrying out the analysis [4]. The demand for proof of competence from regulatory bodies and customers means that PT is relevant to all laboratories testing food and feed for quality and safety in every country. Hence, it is a requirement of accreditation to International Standards Organization (ISO) 17025 [5] that the laboratory takes part in proficiency testing, if a suitable scheme exists. Further, for laboratories entrusted with the official control of food and feeds, Article 12 of EU Regulation (EC) 882/2004 [6] requires such laboratories to be assessed and accredited in accordance with ISO 17025, so that PT is a legal

requirement for these laboratories. Thus, together with the use of validated methods, proficiency testing is an essential element of laboratory quality assurance.

Proficiency testing—the process

FAPAS® has been organising proficiency testing for chemical analysis in food since 1990 in accordance with the International Harmonised Protocol for the Proficiency Testing of (Chemical) Analytical Laboratories [4]. The FAPAS® PT scheme is accredited by the UK Accreditation Service in accordance with the requirements of International Standard ISO/IEC Guide 43-1:2007 through assessment against International Laboratory Accreditation Cooperation Guidelines G13:2007 and relevant elements of ISO 9000:1994. More than 2,000 laboratories in approximately 100 countries take part in a wide range of FAPAS® proficiency tests for analytes such as additives, contaminants, nutrients, residues and allergens in foods. Test

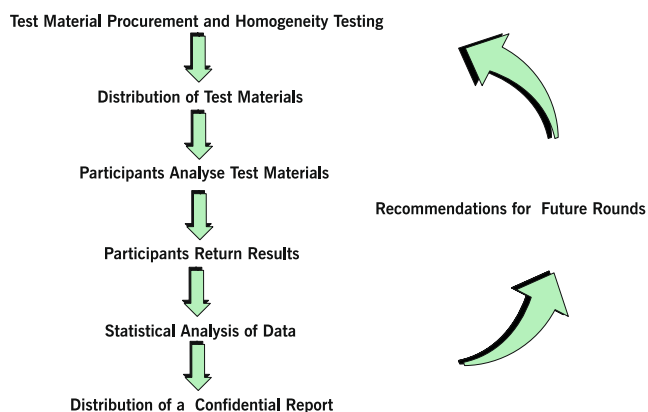


Fig. 1 A flow diagram of the PT process

materials are selected and analysed to confirm sufficient homogeneity prior to distribution to registered participants. Laboratories usually have 6 to 8 weeks to analyse the samples and submit their results. Statistical analysis is carried out on the submitted results data set and a report compiled for issue to participating laboratories. The identity of laboratories taking part in FAPAS® PT remains confidential at all times, although the anonymous scores and reported methods of all participants in each test are included in the final report for comparison purposes. Figure 1 indicates the process by which FAPAS® proficiency testing is conducted.

Test materials

A range of matrix/allergen combinations are available for proficiency testing. In addition to the well-established allergens that cause immune-mediated responses, the list includes foods which can cause non-immune mediated reactions and are associated with food intolerance rather than allergy (e.g. histamine in canned fish, sulphites in jam). As an example, Table 2 shows the proficiency tests available during 2008/2009.

The ultimate effectiveness of a proficiency test depends on the quality of the test materials distributed. There are two major criteria that must be met to produce a suitable test material:

1. *Similarity to samples analysed routinely by participants:* test materials should resemble real samples as closely as possible, which ideally should be food samples naturally incurred with the analyte of interest. The test materials for allergen analysis are always prepared from retail samples purchased locally, by milling and tumbling prior to weighing into pouches which are then sealed and stored in appropriate conditions. For qualitative tests, two test materials are supplied, one

which is positive for the allergen and the other which is allergen-free. The level of allergen in the test materials is controlled by mixing samples known to contain the allergens of interest with allergen-free matrix. Histamine and sulphites are added to the matrix as a chemical spike, and a single test material is supplied for quantitative analysis.

2. *Homogeneity of the bulk test material:* meaningful comparison of participants' results can only take place if each participant receives an identical test material. The International Harmonised Protocol [4] stipulates a strict homogeneity testing procedure by which randomly selected units from a fully homogenised bulk test material are tested for each relevant analyte. Ten samples are analysed in duplicate in random order and statistical analysis of the data set assesses the test material for sufficient homogeneity [7]. The data must pass the appropriate statistical tests which demonstrate that the test materials are not heterogeneous before they are distributed to participating laboratories. Homogeneity data are not included in subsequent calculations of any assigned value. Quantitative analysis for homogeneity of the positive test materials is carried out by a competent laboratory, and five randomly selected blank test materials are analysed to confirm absence of the allergen if appropriate.

Analysis

Participating laboratories are allowed 6 to 8 weeks to carry out their analysis and submit results by the specified closing date. They are encouraged to use the analytical method of their own choice and are asked to provide information on their methodology and the kit used. An example of instructions sent to participants with the test material is given in Table 3.

Data assessment

For quantitative results, each laboratory's performance is assessed by assigning a z-score, which is defined as:

$$z = (x - \hat{X}) / \sigma$$

Where

- x is the measurement of analyte concentration in the test material
 $\hat{X}$ is the assigned value, and
 σ is the target value for standard deviation.

Table 2 Proficiency tests for allergens 2008/2009

Matrix	Allergen	Comments
Chocolate	Hazelnut, hazelnut protein Peanut, peanut protein	Report qualitative and/or quantitative results
Canned fish	Histamine	Histamine added to fish as a spike. Quantitative results required (mg histamine/kg fish).
Cake mix	Gluten Egg, egg white protein	Report qualitative and/or quantitative results
Jam	Sulphites	Sulphite added to jam as sodium metabisulphite. Quantitative results required (mg SO ₂ /kg jam)
Infant soya formula	Casein Beta-lactoglobulin	Report qualitative and/or quantitative results
Infant formula	Lactose	Report qualitative and/or quantitative results
Infant cereal	Gluten	Report qualitative and/or quantitative results
Infant milk formula	Soya, soya protein	Report qualitative and/or quantitative results

The assigned value, $\hat{X}$, i.e. the best estimate of the true concentration of each analyte, is usually obtained from the robust mean [8] of the entire data population for the analyte in question. The procedure used to derive this consensus may involve some or all of the following steps:

- Removing non-valid data, for example non-numerical results, semi-quantitative results
- Scrutinising the results with a novel procedure to identify multimodality ('bump-hunting') [9]
- Considering the normality (Kolmogorov–Smirnov test), or otherwise, of the distribution of the selected results
- Assessing the uncertainty (u) of the mode(s) and the robust mean. For the mode(s), u is taken to be directly equivalent to the standard error of the mode. For the robust mean:

$$u = \frac{\hat{\sigma}}{\sqrt{n}}$$

where

- $\hat{\sigma}$ the robust standard deviation [8] (NB this is *not* the target standard deviation for the round (σ_p) and
- n the number of data points used to calculate the robust mean.

The target standard deviation (σ_p) determines the limits of satisfactory performance in a FAPAS® proficiency test. It is set at a value that reflects best practice for the analysis in question and is ideally prescribed by legislation. In the absence of legislation, the standard deviation of reproducibility found in collaborative trials is generally considered an appropriate indicator of the best agreement that can be obtained between laboratories, although not all analyses have been characterised in this manner. In such cases, the predictive models of the appropriate form of the Horwitz equation [10] are valuable indicators of best practice.

However, FAPAS® has run 45 proficiency tests for allergens over the past 7 years, and it has been clearly

Table 3 Example of instructions sent to participants with their test material

Instructions:
1) Treat the test material as if it was a sample for routine analysis, i.e. you may use any method of analysis you wish.
2) You are required to determine the presence of and/or quantify the level of <i>hazelnut/hazelnut protein</i> present in the test material. Please note: Please indicate your limit of detection (LOD) for all results
It is important that the <i>full kit name</i> is recorded in the appropriate boxes
<i>Qualitative results</i>
Please indicate whether hazelnut/hazelnut protein was 'detected' or 'not detected' in the test materials
<i>Quantitative results</i>
The level of hazelnut or hazelnut protein, in the test materials, should be expressed in <i>mg/kg, as received</i> . If you cannot detect the presence of hazelnut or hazelnut protein in one or both test materials, please state 'not detected' (nd). If you can detect but not quantify the level of hazelnut or hazelnut protein in one or both test materials, please state 'not quantifiable' (nq)
It is important that you report the results in this way so that we can include as many results as possible in the statistical analysis. You are reminded that the ability to report results in the specified units and within the given time scale are part of the proficiency test

Table 4 Quantitative results for determination of hazelnut in chocolate

Analyte	ELISA kit used	Assigned value (mg/kg)	Reported range (mg/kg)	Total number of results	% satisfactory (i.e. $ z \leq 2$)
Hazelnut	R-Biopharm	26.1	14.6–528	17	94
Hazelnut	‘Others’ (3 manufacturers)	N/a	13.6–26.5	5	N/a
Hazelnut protein	ELISA Systems	1.9	1.5–2.6	6	100
Hazelnut protein	Tepnel	3.1	1.73–3.5	5	100
Hazelnut protein	‘Others’ (2 manufacturers)	N/a	2.4–517.4	9	N/a
Hazelnut protein	None-ELISA method	N/a	N/a	1	N/a

demonstrated that use of the Horwitz equation to estimate the standard deviation for proficiency of enzyme-linked immunosorbent assay (ELISA)-based methods is not appropriate. Hence, in the absence of relevant collaborative trial data, the target standard deviation is derived from a value for RSD_R of 25%. Previous experience of analysing data from ELISA methods indicated that this was fit for purpose. The FAPAS® Advisory Committee agreed, and the decision was endorsed in practice by the outcomes of the allergens proficiency tests. Qualitative data are assessed by comparison with the consensus of submitted results and are indicated to participants as either agreeing (A) or disagreeing (D) with that consensus.

Performance assessment

For quantitative data, laboratories are assigned a z -score for each result they submit, and a confidential report for each participant is prepared. z -Score markings are shown in both tabular and ordered histogram form. Since each laboratory is free to choose their own method of analysis, thus simulating the testing of a routine laboratory sample, it is possible to compare the methods used and performance. Whilst the reported result may be dependent upon the method chosen, it could also be affected by the performance of the laboratory equipment or the competence of the analyst. It is therefore important to understand the limitations of this external means of quality assessment when assessing the competence of a laboratory. If it is assumed that the assigned value ($\hat{X}$) and the target standard deviation (σ_p) correctly describe the variation of normally distributed participants’ results (x), the data are known as “well behaved”. In such cases, there is a one in 20 chance that a participant’s well-behaved result may have a z -score of $|z|$ between 2 and 3 (questionable result) and about a one in 300 chance that it may have a z -score of $|z| > 3$ (unsatisfactory result). Detailed information and worked examples of statistical procedures are given in the FAPAS® Protocol [11].

Proficiency testing for allergens—two case studies

Detection assays are available for many of the 14 allergenic ingredients listed in Table 1, including the most prevalent allergens such as peanut, hazelnut, gluten, egg and sesame. A comprehensive review of methods for allergen analysis was undertaken by Poms et al. [12], who emphasise that reliable detection and quantification methods are essential for compliance with food labelling regulations. Most allergens are proteins, which are not always well-characterised chemically, and most methods detect the allergen itself or specific marker compounds (proteins or DNA). Many allergen-specific kits are available, and the most commonly used is the enzyme-linked immunosorbent assay, since it is precise and reasonably simple to use. Obviously limits of detection are an important factor in measuring trace levels of allergens, and Poms and Anklam [13] reported that this should be in the range of 1 to 100 mg/kg, depending on the food and allergen. Limits of detection (LODs) are generally reported between <1 to 5 mg allergen/kg food matrix, although much higher LODs have been noted for some allergens and kits. A major problem encountered in testing the proficiency of quantitative allergy analysis is that the results submitted by participating laboratories can fall into more than a single data set. As described above, the assigned value for a chemical analysis proficiency test is usually obtained from the robust mean of the entire data population for the analyte in question. On occasions, multimodality [9] is noted in the results set from a single test, frequently associated with the method used. However, for allergen testing, it has become apparent that, for a number of reasons, the use of a specific manufacturer’s kit may be associated with an individual mode in the total data set. For instance, the extraction buffer is specific to each manufacturer’s kit and may extract a different fraction of the protein. In addition, each method supplies its own standard calibrators, since food allergens such as peanut, hazelnut or sesame do not have a single recognized standard reference material. For example, peanut kits from different manufacturers are often not

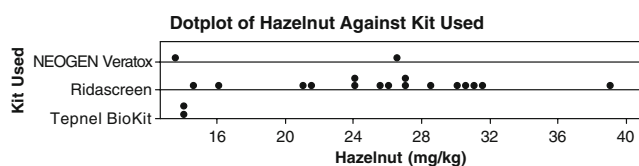


Fig. 2 Dot plot of hazelnut data for chocolate test material by kit brand

calibrated against the same thing. They could be calibrated against different varieties of peanut, which have different protein profiles. They could be calibrated against peanut that has been processed differently, such as roasted or raw. The amount of roasting time will affect the protein profile, so two batches of roasted peanut can still be quite different. Each method will still express results as parts per million (ppm) of peanut or ppm of peanut protein, but the “ppm of peanut” from two different methods does not refer to exactly the same thing, so that the reported results for the allergen and the protein may form individual populations, and statistical assessments must be carried out on each sub-set. Unfortunately, this segregation can mean that some of the resultant data sets are too small to permit a meaningful statistical evaluation. It is important to realise that this does not negate the accuracy and/or validity of these results—it is simply that there are insufficient data points for an independent statistical analysis. A good example of this can be seen in a recent proficiency test for hazelnut in chocolate [14].

Case study 1: hazelnut in chocolate proficiency test

Each participant received two chocolate powder test materials, one contaminated with hazelnut powder to give a final concentration of approximately 25 mg hazelnut powder/kg chocolate. The second test material was a pure chocolate blank. Homogeneity testing confirmed that the contaminated chocolate contained approximately 1.67 mg hazelnut protein/kg chocolate and that hazelnut protein was not detected in the pure chocolate blank test material. Results were submitted by a total of 37 participants, 17 of whom reported results for hazelnut using R-Biopharm kits, six reporting hazelnut protein by ELISA Systems and five reporting hazelnut protein using Tepnel kits. There were insufficient data submitted for hazelnut and hazelnut

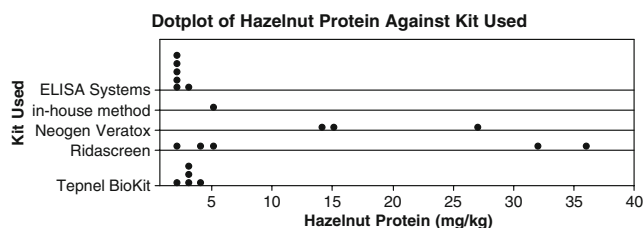


Fig. 3 Dot plot of hazelnut protein data for chocolate test material by kit brand

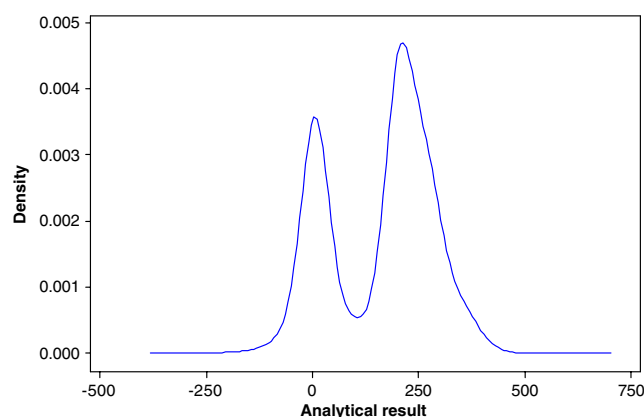


Fig. 4 Kernel density plot of all valid data for sesame in infant cereal (mg/kg)

protein by the remaining nine participants, who used other brands of ELISA kits or non-ELISA-based methods to calculate assigned values. Only one laboratory indicated that they did not use an ELISA method (details of the method were not given). Table 4 gives details of the assigned values and performance. All participants correctly reported the pure chocolate blank test material as negative for hazelnut and hazelnut protein.

Figures 2 and 3 are dot plots showing how results for both hazelnut and hazelnut protein from each kit manufacturer are clustered within the range of reported results. It can be seen that, although there are several results for hazelnut protein by ‘other’ manufacturers’ kits, they are so widespread that the uncertainty of any assigned value calculated would be too great to be meaningful.

The assigned value for total hazelnut (26.1 mg/kg) is close to the concentration of hazelnut powder (approximately 25 mg/kg) added to the pure chocolate. However, two assigned values were derived for hazelnut protein, one from the Tepnel data and the other from the ELISA Systems data. This decision as to whether they are both fit for purpose is for the analytical laboratories to determine but does raise the question as to how to interpret these results in the light of labelling requirements.

Case study 2: sesame in infant cereal

This case study clearly demonstrates how proficiency testing can highlight the very different results that can be obtained from different kits. In October 2006, 26 laboratories submitted results for analysing sesame in infant cereal

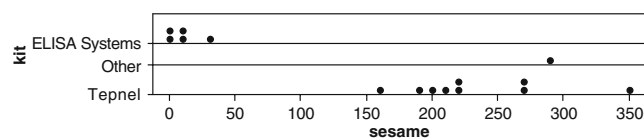


Fig. 5 Dot plot of submitted sesame results by kit brand

[15]. The test material was prepared by milling and mixing together sesame seeds and infant porridge. Results were initially assessed as a single data set but gave a multimodal distribution and were therefore separated into individual kit brands for separate assessment where possible. Seven participants used ELISA Systems kits, of whom five submitted quantitative results, which was too few for independent assessment. There were 12 participants using Tepnel kits, of whom nine submitted quantitative results. Seven results (one quantitative) were submitted by laboratories using a variety of other kits and methods, including in-house and polymerase chain reaction methods. A kernel density plot of all quantitative data (Fig. 4) clearly shows that there are two major modes in the total data set. The lower mode lies within a range of 1.8–26 mg/kg and the major mode between 155–247 mg/kg. Figure 5 shows the distribution of data points for ELISA Systems and Tepnel results, which clearly correlates with the modes in Fig. 4. The median of the range for the major mode was used to set an assigned value (215 mg/kg) for the Tepnel results. The reason for the difference in quantitative values reported for each kit is unclear, as the instructions for the test asked participants to report total sesame measured in the infant cereal.

Conclusions

It is apparent from the two case studies that different brands of test kit can give different populations of results. Consequently, proficiency testing has to provide z-scores for individual kits and methods, providing there are sufficient participants in each population. It is usually impossible to treat all the data submitted for an allergen test as a single population. There may be more than one valid target analyte, for instance either the total content or the allergen protein content, as seen with hazelnut test. Even

where it appears that the analyte is clearly defined, as for sesame, the proficiency test emphasises the variety of data sets that are generated by different methods and kits. This results in a comparison of the method/kit performance rather than laboratory performance. The proficiency test does not judge the kits and methods as to their ability to determine a result close to a 'true' value, since the assigned values are derived as a consensus of each data set. The value of allergen proficiency tests is to provide laboratories with the tools to assess whether their methods are fit for the purpose of assessing allergen content of food.

References

1. Tariq SM, Stevens M, Matthews SM, Ridout S, Twiselton R, Hide DW (1996) *Br Med J* 313:514–517
2. Hourihane JO'B, Kilburn SA, Nordlee JA, Hefle SL, Taylor SL, Warner JO (2007) *J Allergy Clin Immunol* 119:1197–1202
3. Request No. EFSA-Q-2003-016: Opinion of the Scientific Panel on Dietetic Products, Nutrition and Allergies on a request from the Commission relating to the evaluation of allergenic foods for labelling purposes (adopted on 19 February 2004), http://www.efsa.eu.int/science/nda/nda_opinions/catindex_en.html
4. Thompson M, Ellison S, Wood R (2006) *Pure Appl Chem* 78 (1):145–296
5. ISO/IEC 17025:2005. General requirements for the competence of testing and calibration laboratories
6. Regulation EC 882/2004, Official Journal L 165:0001-0141
7. Fearn T, Thompson M (2001) *Analyst* 126:1414–1417
8. Analytical Methods Committee (1989) 114:1693–1697
9. Lowthian PJ, Thompson M (2002) *Analyst* 114:1359–1364
10. Thompson M (2000) *Analyst* 125:385–386
11. FAPAS® 2002 Protocol for the organisation and analysis of data, 6th edn.
12. Poms RE, Klein CL, Anklam E (2004) *Food Addit Contam* 21 (1):1–31
13. Poms RE, Anklam E (2004) *GIT Lab J Eur* 1:43–46
14. FAPAS® Allergens Report 2745, October 2008
15. FAPAS® Allergens Report 2727, November 2006