PERSPECTIVE

Recommendations for the design of laboratory studies on non-target arthropods for risk assessment of genetically engineered plants

Jörg Romeis · Richard L. Hellmich · Marco P. Candolfi · Keri Carstens · Adinda De Schrijver · Angharad M. R. Gatehouse · Rod A. Herman · Joseph E. Huesing · Morven A. McLean · Alan Raybould · Anthony M. Shelton · Annabel Waggoner

Abstract This paper provides recommendations on experimental design for early-tier laboratory studies used in risk assessments to evaluate potential adverse impacts of arthropod-resistant genetically engineered (GE) plants on non-target arthropods (NTAs). While we rely heavily on the currently used proteins from *Bacillus thuringiensis* (*Bt*) in this discussion, the concepts apply to other arthropod-active proteins.

A risk may exist if the newly acquired trait of the GE plant has adverse effects on NTAs when they are exposed to the arthropod-active protein. Typically, the risk assessment follows a tiered approach that starts with laboratory studies under worst-case exposure conditions; such studies have a high ability to detect adverse effects on non-target species. Clear guidance on how such data are produced in laboratory studies assists the product developers and risk assessors. The studies should be reproducible and test clearly defined risk hypotheses. These properties

Handling editor: Paul Christou.

J. Romeis (✉)
Agroscope Reckenholz-Tänikon Research Station ART,
Reckenholzstr. 191, 8046 Zurich, Switzerland
e-mail: joerg.romeis@art.admin.ch

R. L. Hellmich
USDA-ARS, Corn Insects and Crop Genetics
Research Unit, Genetics Laboratory c/o Insectary,
Department of Entomology, Iowa State University,
Ames, IA 50011-3140, USA

M. P. Candolfi
Innovative Environmental Services (IES) Ltd,
Benkenstr. 260, 4108 Witterswil, Switzerland

K. Carstens
Regulatory Science, Pioneer Hi-Bred,
2450 SE Oak Tree Ct., Ankeny, IA 50021, USA

A. De Schrijver
Division Biosafety and Biotechnology,
Scientific Institute of Public Health, Juliette
Wytsmanstraat 14, 1050 Brussels, Belgium

A. M. R. Gatehouse
School of Biology, Institute for Research
and Sustainability, Newcastle University,
Devonshire Building, Newcastle upon Tyne
NE1 7RU, UK

R. A. Herman
Dow AgroSciences LLC, 9330 Zionsville Road,
Indianapolis, IN 46268, USA

J. E. Huesing
Monsanto Company, 800 North Lindbergh Blvd.,
St Louis, MO 63167, USA

M. A. McLean
Center for Environmental Risk Assessment,
ILSI Research Foundation, 1156 Fifteenth St., N.W.,
2nd floor, Washington, DC 20005-17432, USA

A. Raybould
Syngenta, Jealott's Hill International Research Centre,
Bracknell RG42 6 EY, Berkshire, UK

contribute to the robustness of, and confidence in, environmental risk assessments for GE plants. Data from NTA studies, collected during the analysis phase of an environmental risk assessment, are critical to the outcome of the assessment and ultimately the decision taken by regulatory authorities on the release of a GE plant. Confidence in the results of early-tier laboratory studies is a precondition for the acceptance of data across regulatory jurisdictions and should encourage agencies to share useful information and thus avoid redundant testing.

Introduction

Genetically engineered (GE) plants and food and feed products derived from them are strictly regulated by governments internationally. Through the implementation of regulatory systems, designated authorities mandate a pre-market environmental risk assessment and a food/feed safety assessment of GE events case-by-case. These evaluations are a prerequisite to the regulatory decision to approve or not approve GE events for cultivation and for human food and/or livestock feed consumption.

Environmental risk assessment of GE plants is designed to answer very specific questions about the potential risks of introducing such plants into the environment, and includes three main phases: problem formulation, analysis (data collection), and risk characterization (USEPA 1998; Raybould 2006; Carstens et al. 2010; Wolt et al. 2010). Identification of protection goals (e.g., the protection of valued arthropods) is a crucial early step in problem formulation. Protection goals reflect the social, cultural, economic, environmental, and related development objectives of

a country, and therefore are typically specific to each regulatory system (Raybould and Quemada 2010). However, among different regulatory authorities there are some common areas of concern. One of these concerns, and the subject of this paper, is potential adverse impacts that a GE plant may have on non-target arthropods (NTAs),[1] and in particular adverse effects that may arise through exposure to toxins produced in arthropod-resistant GE crops. The magnitude of risk to NTAs depends on the likelihood and seriousness of harmful effects that may result from cultivation of the crop. Generation of relevant effects and exposure data for such toxins is fundamental for any assessment of impacts on NTAs.

Testing of hypotheses about the likelihood of harmful effects of cultivating the GE plant using data collected in the analysis phase is crucial to the outcome of the risk assessment and ultimately the decision taken by regulatory authorities on the release of a GE plant. Useful data of sufficient quality may already exist in the scientific literature or previously conducted studies, or may be acquired from new studies carried out especially for the risk assessment. WHO (2008) described four qualitative criteria that indicate data quality in studies used for chemical exposure assessment:

1. *Appropriateness*: The degree to which data are relevant and applicable to a particular exposure assessment.

2. *Accuracy*: The degree to which measured, calculated, or modeled values correspond to the true values of what they are intended to represent.

3. *Integrity*: The degree to which the data collected and reported are what they purport to be.

4. *Transparency*: The clarity and completeness with which all key data, methods, and processes, as well as the underlying assumptions and limitations, are documented and available.

A. M. Shelton
Cornell University/NYSAES, 630 W. North St., Geneva,
NY 14456, USA

A. Waggoner
Office of Pesticide Programs, Biopesticides and Pollution
Prevention Division, US Environmental Protection
Agency, 1200 Pennsylvania Avenue, NW, Washington,
DC 20460, USA

---

[1] As current and near future commercial products target arthropod pests, this paper focuses more narrowly on direct effects on this group of animals as a sub-group of non-target organisms (NTOs). A further restriction is that non-target arthropods (NTAs) are generally considered beneficial as opposed to non-target pests. The concepts discussed here will apply to other non-target invertebrates, e.g., earthworms and nematodes.

Table 1 Ranking of types of information and their relative values as evidence (OGTR 2009) (GMO refers to the genetically engineered organism)

	Reliability	Appropriateness
↑ Increasing value	Validated studies conducted according to international protocols meeting defined standards	Experimental data on the GMO and/or parent organism in the Australian environment
	Peer reviewed literature—strongly supported reports, models, theories	Experimental data on the GMO and/or parent organism overseas
	Peer reviewed literature—single report, model, theory	Experimental data on modified traits in other organisms
	General biological principles	Experimental data on related, surrogate systems
	Opinion of an expert familiar with the GMO, parent organism, modified traits, ecology	
	Other technical reports, specialist literature (for example, beekeeping), government reports, etc.	
	No information to indicate a problem	
	Unsubstantiated statements	

The importance and applicability of these "hallmarks of data quality" (WHO 2008) to the environmental risk assessment of GE plants has been recognized implicitly or explicitly by regulatory authorities who have attempted to define mutually acceptable quality standards for chemical exposure in regulatory submissions. These same standards should also apply to GE plants. For example, the Risk Analysis Framework published by the Australian Office of the Gene Technology Regulator (OGTR 2009) addresses the quality of evidence (Table 1). The provision of such guidance about data quality assists both the product developer and strengthens the robustness of the risk assessment.

This paper summarizes the outcomes of expert panels convened by the West Palaearctic Regional Section of the International Organisation for Biological Control (IOBC/WPRS) in 2007 and the International Life Sciences Institute (ILSI) Research Foundation in 2009, with additional contributions from subject matter experts. It provides guidance and recommendations on experimental design for early-tier laboratory studies (termed Tier I and/or Tier II studies depending on the jurisdiction) used to evaluate potential adverse effects of arthropod-resistant GE plants on NTAs. The specific recommendations provided herein may be viewed as the basic quality standard for early-tier NTA studies used to support regulatory submissions.

Assessment of GE plant effects on non-target arthropods

Problem formulation directs the scope of risk assessment and defines the environmental entities that are to be protected (termed assessment endpoints) against a potential stressor. For example, beneficial arthropods are valued ecological entities; abundance within the agroecosystem is an important attribute; thus "beneficial arthropod abundance" constitutes an assessment endpoint. Problem formulation further generates testable scientific hypotheses and endpoints to measure (termed measurement endpoints) that are relevant for decision-making and are subsequently addressed in the analytical phase of the risk assessment (USEPA 1998; Raybould 2006; Carstens et al. 2010; Wolt et al. 2010). Problem formulation should culminate in a conceptual model delineating how harm can occur by a particular stressor on the assessment endpoint (including an analysis of whether or not exposure to the stressor occurs), leading to an analysis plan that is consistent with the risk hypotheses and should establish the relationship between the stressor and the ecological impacts of concern. A typical risk hypothesis related to NTA effects of an arthropod-resistant GE plant is: "The expressed protein[2] is not toxic to NTAs at the

---

[2] Although insecticidal traits associated with commercialized GM crops have all been based on proteins, we recognize that

Fig. 1 Risk assessment continuum. The tiered risk assessment moves from tests that have a high ability to assess adverse effects to more complex experiments under field conditions that evaluate the risks under more realistic exposure conditions. Power refers to the ability to evaluate adverse effects
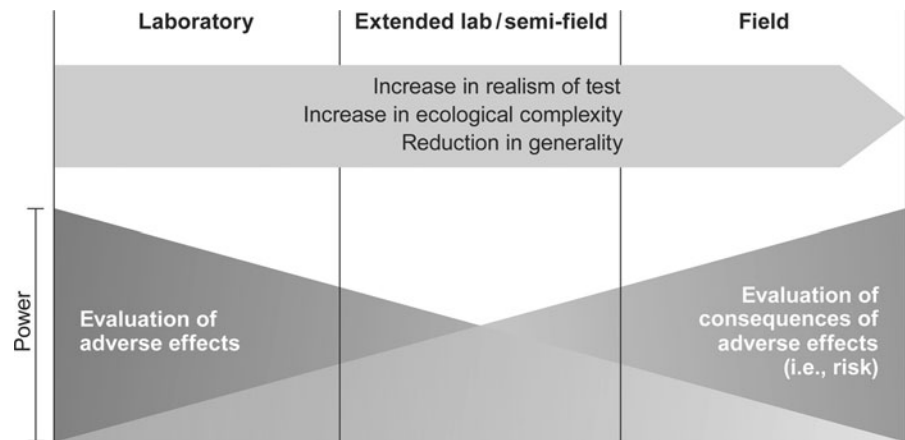
concentration present in the field" (Raybould 2007; Romeis et al. 2008).[3]

The risk hypotheses are then addressed in the analysis phase of the risk assessment following a tiered framework that is conceptually similar to that used to assess the environmental impact of conventional chemical plant protection products (Hill and Sendashonga 2003; Garcia-Alonso et al. 2006; Rose 2007; Romeis et al. 2008). Based on the risk hypotheses, early-tier laboratory experiments are conducted under worst-case exposure conditions where species representative of NTAs present in the receiving environment that are likely to be exposed to the arthropod-active protein [referred to as surrogate species by Caro and O'Doherty (1999)] are exposed to concentrations of the protein in excess of exposure in the field. This increases the likelihood of detecting adverse effects on NTAs (Fig. 1).

Protocols developed to assess the impact of chemical plant protection products on NTAs have provided a useful basis for designing similar protocols to assess the potential effects of GE plants on NTAs (Romeis et al. 2008). They indicate which species may be suitable surrogates for laboratory studies, describe general procedures including test system description, organism preparation, test diets, experimental design as well as suitable measurement endpoints and quality criteria such as acceptable control mortalities to adequately address the assessment endpoint. Available protocols range between statements of general principles (e.g., USEPA 1996a, b) and species specific guidance documents (e.g., Candolfi et al. 2000; Grimm et al. 2002). Many of these protocols have been modified to consider the oral exposure pathway of plant-expressed arthropod-active proteins, and several protocols of this type have been described in the literature (e.g., Stacey et al. 2006; Duan et al. 2002, 2006, 2007).

In some regulatory jurisdictions, early-tier NTA studies may have to be conducted under the defined standards of Good Laboratory Practice[4] (GLP) which includes study reconstructability (OECD 1998). The particular requirements of GLP studies are outside the scope of this paper and it is advisable for scientists conducting NTA studies as part of regulatory submissions to ensure that GLP is followed if and as required by regulatory authorities.

If no adverse effects are seen under the worst-case exposure conditions in early-tier laboratory studies,

---

Footnote 2 continued

future traits will not be restricted to this class of molecule. For simplicity we restrict ourselves to arthropod-active proteins. However, most of the considerations that are described will be equally applicable for compounds belonging to other classes of molecules.

[3] A second risk hypothesis relevant to NTA risk assessment is that the transgenic crop does not contain any harmful unintended changes that may result from the transformation process. This hypothesis is not addressed in the present paper since it is generally tested using a range of data that establish the equivalence and familiarity of the GM plant with its non-transformed comparators (Romeis et al. 2008; Raybould et al. 2010).

---

[4] Good Laboratory Practice (GLP) embodies a set of principles that provides a framework within which laboratory studies are planned, performed, monitored, recorded, reported and archived. GLP helps assure regulatory authorities that the data submitted are a true reflection of the results obtained during the study and can therefore be relied upon when making risk/safety assessments.

the risk can be characterized as being acceptable and there may be no need to conduct any further testing because of the minimal probability of adverse effects in the field where NTAs are exposed to much lower concentrations of the arthropod-active protein. Early-tier testing thus allows elimination from further consideration risks that are negligible, and allows assessors to focus resources to address more significant risks or uncertainties. If effects are seen under laboratory conditions at high test substance exposure concentrations, the risk can be further characterized in additional laboratory or higher-tier experiments that use more realistic environmental exposure scenarios (Fig. 1). Higher-tier studies can include semi-field tests under enclosed (contained) conditions and open-field tests, and are sometimes conducted when evaluations across multiple trophic levels are warranted or estimation of population parameters is sought. The studies may involve the use of population and community responses and may consider geographic and temporal variability in exposure to the stressor of concern. Higher-tier tests are demanding in terms of skills and resources necessary for their design, execution, and analysis. Furthermore, results that are difficult to interpret often do not contribute additional confidence in the conclusions of the risk assessment. A recent meta-analysis of published studies on non-target effects of *Bt* crops has confirmed that laboratory studies "…predicted effects that were on average either more conservative than or consistent with effects measured in the field" (Duan et al. 2010).

In conclusion, higher-tier tests should only be conducted

1. when triggered by the detection of potentially adverse effects in lower tiers of testing;
2. when tests at lower tiers are not possible (e.g., due to a lack of a testable surrogate or lack of a validated test protocol);
3. when the nature and mode of action (MOA) of the arthropod-active protein being evaluated suggest that the higher-tiered test is most appropriate to detect effects.

Selection of surrogate species and life-stages

Since it is not possible to test all species that are potentially present in the receiving environment and

exposed to the arthropod-active protein, surrogate species should be selected that represent different habitats (e.g., soil- or plant-dwelling arthropods) or different ecosystem services such as ecological functions (e.g., predator, parasitoid or decomposer), and taxonomic groups. To test the risk hypotheses that were generated in the problem formulation phase, the subset of species and life-stages selected for early-tier testing should be chosen based on the potential exposure pathway, knowledge on the spectrum of activity and the MOA of the arthropod-active protein, the amenability of the test system[5] for the selected NTA, and the availability of the test organism:

1. *Exposure pathway.* Surrogate species that are tested should be representatives of those that are most likely to be exposed to the arthropod-active protein in the field. There may be considerable knowledge available on the fate of the protein (e.g., *Bt* Cry protein) under field conditions and its movement through arthropod food-webs (Romeis et al. 2009). Consequently, experiments may not need to be conducted to inform the risk assessment for species and life-stages that are at negligible risk because of limited exposure. In the case of *Bt*-transgenic crops, an example would be predators and parasitoids that specifically attack aphids which are known to contain no or only trace amounts of *Bt* Cry protein (Romeis and Meissle 2010). Another example is the negligible exposure of pollinators or pollen feeders where there is a lack of protein-expression in the pollen, for example when the transgene is controlled by a promoter that does not yield expression in pollen.

2. *Known spectrum of activity and mode of action of the arthropod-active protein.* Surrogates and life-stages need to be selected that are most likely to be susceptible to the arthropod-active protein and thus are most likely to detect an adverse effect (i.e., have the highest predictive power). For example, in the case of

---

[5] OECD (2005) defines "test system" as "any biological, chemical or physical system or a combination thereof used in a study." US EPA 40CFR160 (USEPA 2009) defines "test system" as "any animal, plant, microorganism, chemical or physical matrix, including but not limited to soil or water, or subparts thereof, to which the test, control, or reference substance is administered or added for study". Test system also includes appropriate groups or components of the system not treated with the test, control, or reference substance.

current commercialized *Bt* Cry toxins, immature holo-metabolous insects are the only arthropods showing meaningful susceptibility and neonates are more sensitive than later instars (e.g., Glare and O'Callaghan 2000). When the arthropod-active protein is known to affect immature development or fecundity in sensitive arthropods (e.g., target pests), testing of multiple life-stages or adults, respectively, would be appropriate.

3. *Amenability to testing*. Species and life-stages should be selected that are amenable to testing under laboratory conditions (Rose 2007; Romeis et al. 2008). This includes the availability of experimental protocols that ensure the interpretability of the data and the possibility to reproduce or reconstruct the study. Experience with the testability of a species or a specific life-stage comes, for example, from NTA testing of pesticides and from previous studies with orally active proteins. Different life-stages of a particular arthropod species (e.g., Meissle and Romeis 2009b) or of taxonomically-related species (e.g., Heimbach et al. 2000a) might differ substantially in their amenability to testing in the laboratory due to their specific biology, lifespan, required diet, etc.

4. *Availability of the test arthropod*. For pesticide testing, the IOBC/WPRS, European Plant Protection Organisation (EPPO) and Beneficial Arthropod Regulatory Testing Group (BART) have recommended using laboratory-reared arthropods (Barrett et al. 1994; Candolfi et al. 2000, 2001). Standardized test arthropods from public sector or commercially reared laboratory colonies provide a level of consistency between experiments (due to an overall similar genetic susceptibility) and testing laboratories that promotes data reproducibility and comparability. Although some phenotypic differences from wild populations may occur during laboratory breeding, such limitations are deemed preferable to the unknown and variable condition of field-collected specimens (e.g., previous exposure to the test substance, variable age and health status). Although not recommended, when there is not a viable alternative and field-collected NTAs must be used, specimens should be standardized as much as possible and information on the site and method of collection as well as details on the handling and maintenance between the time of collection and use in the experiments should be provided (e.g., Heimbach et al. 2000a, b).

Test substance

The test substance should be characterized and formulated in a way that allows precise calculation of the amount that is delivered to the test organism.

Test substance characterization and equivalence

Relevant properties of the test substance such as physical state, color, consistency, and pH should be described. Before starting an experiment, the following should be known about the test substance, which is typically either in a purified form or GE plant material expressing the protein of interest:

1. *Biological activity of the test substance.* Test substances should be in a formulation that ensures biological activity. Bioactivity should be confirmed using relevant assays like sensitive insect bioassays (e.g., Duan et al. 2006, 2008a; Stacey et al. 2006; Meissle and Romeis 2009a), biochemical assays in case of enzyme inhibitors (e.g., Shade et al. 1994), or agglutination assays in case of lectins (e.g., Van Damme et al. 1987).

2. *Purity of the test substance.* The purity of the test substance needs to be known to calculate the true dose and amount that is delivered to the test organism. The active ingredient(s) and relevant impurities must be identified and quantified within technically feasible limits and the method(s) applied to determine purity must be stated.

3. *Test substance equivalence* (*e.g., for purified protein*). The arthropod-active protein provided to test organisms may be derived from GE plants or, more commonly, produced as recombinant protein in microbial expression systems. The concentration of the introduced protein in transgenic plant tissues can be very low, often less than 0.01% on a dry weight basis. Early-tier studies (and other toxicology studies such as those used to assess GE food safety) that require relatively large amounts of test substance are often not feasible using plant-expressed protein as sufficient mass cannot be reasonably purified from the plant source (CAC 2003). Instead, these studies normally make use of protein purified from bacterial or yeast expression systems. In such cases, it is necessary to demonstrate functional and biochemical equivalence (i.e., equivalent physiochemical properties and biological activities) of

the plant and microbial purified proteins (Gao et al. 2004, 2006; Raybould and Vlachos 2010). For example, comparisons of the molecular weight, the isoelectric point, amino-acid sequence, post-translational modifications including glycosylation patterns, immunological reactivity, biological activity, and in the case of enzymes, the enzymatic activity, may be needed to provide evidence for the equivalence (EFSA 2008; USEPA 2000, 2001). For Cry and vegetative insecticidal proteins (VIP) from *Bacillus thuringiensis*, fully characterized purified proteins have been successfully produced for use in NTA studies (Gao et al. 2004, 2006; OECD 2007; Raybould and Vlachos 2010), but this may not always be possible for future arthropod-active proteins.

Test substance stability and homogeneity

To ensure consistent exposure to the test substance over the course of a laboratory study, its stability has to be ensured or the diet into which it has been incorporated has to be replaced from a characterized batch at regular intervals. Where the test substance has been incorporated into a diet, its concentration should be monitored and recorded throughout the test period. Immuno-assays such as enzyme-linked immunosorbent assay (ELISA) to measure protein concentration, Western blot analysis to measure protein intactness, and/or sensitive insect bioassays to measure bioactivity (Duan et al. 2002, 2006, 2008a; Stacey et al. 2006; Raybould and Vlachos 2010) may be employed. Stability should be defined based on the variability of the assay used. For example, for some ELISA methods, stability may be defined as >70% of initial concentration. Stability criteria should be defined prior to study initiation and should consider limitations of the immunoassays, insect bioassays, or effects of reagents and buffers on the test substance. In general, the following principles should be followed:

1. The test substance (or a diet medium into which it has been incorporated) should be stored under conditions that maintain its intactness and activity. Note that freeze–thaw cycles may be detrimental to some substances, especially proteins. If this is the case, very short harvest-to-analysis intervals must be used or other storage conditions must be devised.
2. The batch of the test substance tested should preferably be the same throughout the duration of the experiment.

3. If it is not possible to use the same batch throughout the experiment, a new batch of test substance can be used as long as it is fully characterized.
4. If the stability of the test substance cannot be guaranteed for the duration of the study, freshly treated diet should be supplied periodically (e.g., daily).

Appropriate assays should be performed to ensure the storage stability of the test substance during the experiment. For example, a sub-sample of the test substance should be stored under the same conditions (e.g., sub-freezing temperatures) as the samples that are used in the experiment, and analyzed at the end of the storage period.

If the test or control substance is incorporated into a diet, testing must confirm that the method of mixing results in the expected concentration of the protein and that the test substance is homogenously distributed in the diet. This determination should occur prior to the start of the study or concurrent with it. Homogenous distribution of the test substance is important to rule out that individual test organisms are able to avoid the test substance altogether or are exposed to lower than anticipated levels while others might be over-exposed. Homogeneity of the diet is typically tested by analyzing sub-samples of the diet (e.g., Duan et al. 2006, 2008a). Homogeneity criteria, as with stability criteria, should be defined prior to study initiation and should account for assay limitations. Similarly, it needs to be known to what extent the test substance activity degrades at different times during the experiment in order to calculate the actual dose delivered. One could refer to published studies when available. For example, when *Bt* maize pollen is used as a carrier to expose NTAs to a Cry protein, ELISA analyses may be sufficient to establish the concentration if stability has been previously established under relevant environmental conditions (Wraight et al. 2000; Hellmich et al. 2001; Stanley-Horn et al. 2001; Meissle and Romeis 2009a).

Method of delivery (carrier)

When the test substance is expected to act by a dietary route, the test substance must be delivered orally. To date, this has been true for all plant-expressed arthropod-active proteins. The method of delivery of the test and control substances should be selected to

ensure maximum accuracy of the dose administered. It needs to be established that the chemical and biological properties of the test substance are not altered when incorporated into a carrier (see also above section on "Test substance stability and homogeneity"). In addition, appropriate controls must be added to the study design to differentiate effects that are related to the carrier and those that are related to the test substance (see also below section on "Control substances"). In all cases, potential effects of the carrier must be accounted for in the experiment. Different methods for delivery of test substance can be applied with the following considerations:

1. *Artificial diet.* In most cases the purified protein or GE plant material is provided to the test organism in artificial diet. When purified arthropod-active protein is used as the test substance it needs to be dissolved in water or buffer and the characteristics of the solvent must be provided. The science and technology of arthropod diets can be complicated (Cohen 2004). Since the diet can directly and/or indirectly impact study quality and study results, the diet is a key element to consider in study design. Since most arthropod diets are meridic (only partly defined chemically), care must be taken that the diet constituents themselves do not adversely interact with test substances. Diets may be species-specific, as determined by the nutritional requirements and the feeding habits of the test NTA. The test substance should be incorporated into the diet in a homogenous manner, or used as a diet overlay, permitting the arthropods to ingest the protein during feeding. In any case, a detailed description of such procedures should be provided. Care must be taken to ensure the test substance is not affected when it is incorporated into the diet (e.g., by heat deactivation).

Artificial diets have successfully been used to test the effects of purified arthropod-active proteins on a range of arthropods, including: larvae of *Aleochara bilineata* (Coleoptera: Staphylinidae) (Stacey et al. 2006), larvae of *Poecilus chalcites* (Coleoptera: Carabidae) (Duan et al. 2006), nymphs of *Orius insidiosus* (Heteroptera: Anthocoridae) (Stacey et al. 2006), adult *Chrysoperla carnea* (Neuroptera: Chrysopidae) (Li et al. 2008), and adult and larvae of *Apis mellifera* (Hymenoptera: Apidae) (Malone et al. 1999; Brødsgaard et al. 2003). Particularly noteworthy are current efforts to standardize artificial-diet

based in vitro feeding methods to test *A. mellifera* larvae (Aupinel et al. 2007, 2009; CDPR 2009).

2. *Treatment of non-GE food items.* In cases when no artificial diet is available, alternative ways to dose the test organisms with the purified protein can be applied. For example, the test substance may be dissolved in appropriate surfactants and applied to non-GE plant material. This has been done with leaves to expose foliar-feeding arthropods (Chen et al. 2008). Similarly, non-GE pollen has been treated with solution in which the test protein was dissolved to dose predators such as *Coleomegilla maculata* (Coleoptera: Coccinellidae) (Duan et al. 2002) or *O. insidiosus* (Duan et al. 2007, 2008a), and bees such as *A. mellifera* (see references in Duan et al. 2008b) and *Osmia bicornis* (Hymenoptera: Megachilidae) (Konrad et al. 2008). Alternatively, the predatory beetles *Coccinella septempunctata* (Coleoptera: Coccinellidae) and *Poecilus cupreus* (Coleoptera: Carabidae) have been exposed to test substances by dipping their prey into toxin-containing solution or by injecting their prey with the solution, respectively (Stacey et al. 2006). Another option exists in dosing the test organisms with test substances dissolved in a sugar-rich solution (honey or sucrose). This method is commonly used to dose adults of parasitic Hymenoptera that are known to feed on carbohydrate sources in the field (Romeis et al. 2003a; Bell et al. 2004; Hogervorst et al. 2009) but also for larvae of predatory arthropods including *C. carnea*, *C. septempunctata* and *Adalia bipunctata* (Coleoptera: Coccinellidae) (Hogervorst et al. 2006; Lawo and Romeis 2008; Álvarez-Alfageme et al. 2010).

3. *GE plant material.* GE plant material containing the arthropod-active protein can be used as the test substance when it is regarded as the stressor of concern or in situations where the test organism cannot be fed on an artificial diet or where no purified protein is available. The study conductor must be aware of the fact that there is also the possibility of carrier effects; in other words, there may be differences between the treatment and control unrelated to the arthropod-active protein and related to the effect of different carriers, for example due to different genetics or environments under which the GE plants were grown (see Escher et al. 2000; Wandeler et al. 2002; Jensen et al. 2010; Knecht and Nentwig 2010 for examples).

For example, larvae of the monarch butterfly, *Danaus plexippus* (Lepidoptera: Danaidae), could not be reared reliably to adult on an artificial diet, so milkweed leaf discs (*Asclepias curassavica*) dusted with varying amounts of *Bt* maize pollen grains were used to estimate pollen/Cry protein effects (Hellmich et al. 2001; Dively et al. 2004). Similar studies have been conducted with larvae of other Lepidoptera species (Wraight et al. 2000; Jesse and Obrycki 2002; Shirai and Takahashi 2005; Li et al. 2005).

Bt maize pollen may also be used as a carrier to expose predatory arthropods that readily consume this plant tissue under field and laboratory conditions. Species that have successfully been exposed this way include: larvae and adults of *A. mellifera* (see references in Duan et al. 2008b), *C. maculata* (Coleoptera: Coccinellidae) (Duan et al. 2002; Lundgren and Wiedenmann 2002), *Propylea japonica* (Coleoptera: Coccinellidae) (Bai et al. 2005), adult *C. carnea* (Li et al. 2008), immature and adult *Neoseiulus cucumeris* (Acari: Phytoseiidae) (Obrist et al. 2006c), and juvenile spiders (Araneae) (Ludy and Lang 2006; Meissle and Romeis 2009b). Furthermore, GE plant litter is commonly used to study impacts on soil arthropods that play a role as decomposers such as Collembolans (Yu et al. 1997; Romeis et al. 2003b; Heckmann et al. 2006) and mites (Yu et al. 1997).

The use of GE plant material may prevent testing of the arthropod-active protein under worst-case exposure conditions in excess of that found in plant material since the test arthropods can only consume the concentration of the active protein contained in the plant tissue. However, since the diet of NTAs is unlikely to consist of 100% GE crop tissue in the field, studies using plant tissue can still provide very conservative exposures (see also below section on "Concentration/dose selection").

4. *GE plant-fed herbivores.* Herbivores that have fed on GE plant material and are subsequently shown to contain bioactive toxin may be used as carriers to deliver the arthropod-active protein to entomophagous arthropods (tri-trophic exposure). Great care should be taken to ensure that the herbivores themselves are not adversely affected by the ingested protein so as to avoid the impact of prey/host-quality mediated effects (Romeis et al. 2006; Naranjo 2009). One such carrier that has successfully been used is *Bt* maize-fed spider mites (e.g., Dutton et al. 2002; Li and Romeis 2010;

Álvarez-Alfageme et al. 2008, 2010). These herbivores have been shown to contain high concentrations of *Bt* protein which are similar to concentrations measured in the leaves on which they have fed (Obrist et al. 2006a, c; Torres and Ruberson 2008; Meissle and Romeis 2009b; Li and Romeis 2010; Álvarez-Alfageme et al. 2008, 2010). Furthermore, sensitive insect bioassays have shown that Cry proteins contained in spider mites after feeding on *Bt* maize retain their biological activity (Obrist et al. 2006b; Meissle and Romeis 2009a). The bioactivity of *Bt* Cry proteins following ingestion of *Bt* maize has also been confirmed for larvae of *Ostrinia nubilalis* (Lepidoptera: Crambidae) (Head et al. 2001; Obrist et al. 2006b) and adult *Diabrotica virgifera virgifera* (Coleoptera: Chrysomelidae) (Meissle and Romeis 2009a).

Another potential carrier of the test substance is a strain of the target organism that is resistant to the particular arthropod-active protein. For example, strains of *Plutella xylostella* (Lepidoptera: Plutellidae) and *Helicoverpa armigera* (Lepidoptera: Noctuidae) that are resistant to specific *Bt* Cry proteins have been used to assess toxin effects on predators and parasitoids (Schuler et al. 2003, 2004; Ferry et al. 2006; Chen et al. 2008; Lawo et al. 2010). Recently, Lawo et al. (2010) measured the Cry1Ac concentration in neonate *H. armigera* after 24 h feeding on *Bt* (Cry1Ac) cotton. Larvae from a Cry1Ac-resistant strain contained four times more toxin per gram fresh weight when compared with larvae from a susceptible strain. Resistant larvae in tri-trophic studies can thus be used to expose their natural enemies to higher concentrations of the toxin when compared with the natural situation in the field. However, care must be taken to ensure that the ingested protein in the resistant insects is still active (Chen et al. 2008).

Concentration/dose selection

As part of the environmental risk assessment, an exposure characterization is performed to determine how much of the arthropod-active protein a particular organism might be exposed to in the field under natural conditions (the expected environmental concentration, EEC). In general, secondary exposure of arthropod predators and parasitoids through herbivorous prey or hosts is generally lower than direct exposure of plant-consuming arthropods. Since in the

majority of cases precise estimates of the concentrations of the arthropod-active protein in the natural diet of the NTA are not possible, conservative assumptions must be made. For this purpose, the highest average concentration of protein measured in the plant tissue over the course of growth development is typically taken as the worst-case EEC (e.g., Rose 2007; Raybould et al. 2007; Raybould and Vlachos 2010). Defining the highest average plant-expression level as the EEC adds to the conservatism of the assessment since (a) the NTAs may not exclusively consume GE plant tissue, and (b) plant-produced arthropod-active proteins are usually diluted in the natural food web. This dilution effect has been reported for *Bt* crops (e.g., Harwood et al. 2005; Obrist et al. 2006a; Meissle and Romeis 2009b) but has also become evident from tri-trophic laboratory studies with other arthropod-active proteins (Bell et al. 2003; Christeller et al. 2005). Nevertheless, some cases exist where the EEC has been defined more precisely taking into account knowledge on the NTA's feeding behaviour. For example, consumption of maize pollen has been quantified in some detail for larvae of *A. mellifera* (Babendreier et al. 2004), adult *C. carnea* (Li et al. 2010) and adult *C. maculata* (Lundgren and Wiedenmann 2004; Lundgren et al. 2005). The EEC can thus be based on pollen expression levels (if the arthropod-active protein is present in pollen) since consumption of this plant tissue is the main (if not only) source of exposure to the plant expressed protein.

In practice, NTA studies are often conducted at times when plant expression levels for the event that is going to be commercialized are not fully characterized. Therefore, studies are often conducted at the highest possible concentration of the test substance that can be delivered with the test system. One has to be aware, that these high doses may eventually cause an effect which would then trigger additional studies including, for example, dose–response tests to assess accurately the effect relative to the EEC, which, as mentioned previously, depends on plant-expression levels.

Early-tier tests are often conducted as single dose tests, for instance, at the so-called maximum hazard dose (MHD). The MHD is calculated by multiplying the EEC with a margin of exposure factor (e.g., US EPA suggests a MHD margin of exposure factor of $10\times$). In cases where a high excess dose is not achievable (e.g., because GE plant material is used as

the test substance), the actual maximum dose to which the test organism is exposed should be reported and the reason for the testing dose selection stated. The MHD margin of exposure factor adds certainty to the conclusions drawn from the test and accounts for possible intra- and interspecies variability from the use of a surrogate species. Studies that establish a lack of adverse effects at the MHD level are sufficient to confirm the absence of unacceptable adverse effects, making lower dose testing unnecessary.

For studies establishing an LC_{50}, LD_{50}, EC_{50}, or ED_{50} value the number of doses and test organisms evaluated must be sufficient to determine accurate values, and when necessary or required, the Lowest Observed Adverse Effect Concentration (LOAEC), or No Observed Adverse Effect Concentration (NOAEC). If the LC_{50}, LD_{50}, EC_{50}, or ED_{50} values are greater than the MHD (e.g., $LD_{50} > 10\times$ EEC), then those data are sufficient to inform the risk assessment, and lower dose testing is unnecessary. If the values are less than the MHD used, additional testing with lower doses may be required.

An estimation of risk is often made by comparing the LOAEC or NOEC to the EEC; when the EEC is lower than the LOAEC or NOEC, a conclusion of 'reasonable certainty of no adverse effects' can be made (e.g., Raybould et al. 2007; Raybould and Vlachos 2010).

For studies using GE plant tissues as the test substance, a factor of $1\times$ or less is used and a margin of exposure of greater than one may not be possible due to the limited expression of the arthropod-active protein in the plant. A $1\times$ plant concentration of the protein may thus be most relevant when the actual exposure to the species of interest in the field is known to be much lower that the concentration of the novel protein in the plant tissue, for example, if a test organism is forced to ingest green plant tissue, but in the environment the organism feeds on decomposing senesced tissue. For some arthropod species, exposure to the test substance can be enhanced when test organisms are exclusively fed with plant or arthropod material containing the protein of interest that would otherwise constitute only part of their diet in the field or by using lyophilized plant powder as test substance. Examples include ladybird beetles that use maize pollen as complementary food in the field or generalist predators that occasionally consume spider mites as part of their prey spectrum. Feeding these species with

either a high proportion, or even exclusively with GE maize pollen or GE plant-fed spider mites, would constitute worst-case exposure. Examples are tests in which larvae or adults of *C. maculata* are fed large amounts of GE maize pollen mixed with insect eggs (Duan et al. 2002; Lundgren and Wiedenmann 2002), studies in which *A. mellifera* receive only GE maize pollen (Rose et al. 2007), or the feeding of predominately aphidophagous *A. bipunctata* or *C. carnea* larvae exclusively with spider mites that have fed exclusively on transgenic plant material (Dutton et al. 2002; Álvarez-Alfageme et al. 2010).

Measurement endpoints

Prior to testing, the objectives of the individual study need to be defined, and specific measurement endpoints (also known as measures of effect) described that test the identified risk hypotheses. Appropriate measurement endpoints should be easy to evaluate in the laboratory and likely to indicate the possibility of adverse effects on the abundance of NTAs or other assessment endpoints. Thus, priority should be given to measurement endpoints for which it is clear what change constitutes an adverse effect.

Typical measurement endpoints to address NTA effects of plant protection products and arthropod-active proteins that are expressed in GE plants are mortality (e.g., estimated as LD_{50}), fecundity, development duration, body mass (as a measure of growth), or the percentage of individuals that reach a certain life-stage (e.g., percent adult emergence) (Candolfi et al. 2000; Rose 2007; see Stacey et al. 2006 and Duan et al. 2002, 2006, 2008a for sample protocols). Independent from the measurement endpoints that are selected for a specific study, risk assessors should agree on how to interpret and use these data in the risk assessment. This includes the definition of thresholds that trigger additional testing. Similar to the assessment of pesticides (Candolfi et al. 2000) an effect size of 50% mortality has been defined as the threshold to trigger additional tests for early-tier laboratory studies conducted under worst-case exposure conditions with purified arthropod-active protein or GE crop tissue in the USA (USEPA 1998; Rose 2007). Less than 50% mortality under these conditions of extreme exposure suggest that population effects are likely to be negligible given realistic field exposure scenarios.

Furthermore, once the threshold is defined it should be ensured that the experiment is sufficiently replicated to detect the defined effect size with acceptable statistical power (see also section below on "Statistical considerations").

Determination of the measurement endpoint(s) should consider knowledge about the impact of the arthropod-active protein on the target organisms and its MOA, knowledge about the biology of the selected NTA species and life-stages, and the availability of reliable test protocols. The measurement endpoint(s) selected will affect the duration of the test (see section below on "Test duration"). For example, the current arthropod-active proteins in GE crops (*Bt* Cry and VIP proteins) are lethal to sensitive (target) species. Thus, mortality is one obvious measurement endpoint for laboratory NTA studies. In the case of arthropod-active proteins that are known to cause sublethal effects on sensitive arthropods (such as reducing the fecundity or delaying development), these parameters should receive attention and be measured when assessing the impact on NTAs. This is the case for some inhibitors that affect the arthropod's digestive system and lectins. Several endpoints may need to be measured for arthropod-active proteins for which limited experience regarding their impact on arthropods exists.

Besides the described measurement endpoints, any other sublethal effects that are observed during the experiment (e.g., changes in behavior) should be recorded. Subsequently, risk assessors may agree that the observed effects trigger additional testing. However, observations of apparent sub-lethal effects may need to be interpreted with caution and always compared with the negative control(s), since rearing arthropods on a sub-optimal diet medium may itself cause unforeseen side-effects on their subsequent reproductive vigor.

Test duration

In general, laboratory tests are shorter in duration than semi-field or field studies but are conducted at higher protein doses/concentrations. The duration of a specific laboratory test depends largely on the endpoints that are measured, i.e., the duration of the test must be sufficiently long for the measurement endpoint to respond should the test substance have an adverse

effect. The duration of an experiment is further determined by the selected surrogate and its life-stages, their rate of development under the specific experimental conditions (incl. experimental set-up, environmental conditions), the suitability of the test system, and the characteristics of the test substance.

1. *Measurement endpoint.* The duration required to measure a certain endpoint depends on the endpoint chosen. In general experiments that focus on the organism's developmental stages and measure, for example, mortality or days to adult emergence can vary from 14 to more than 30 days depending on the species, while for example tests measuring the fecundity of *A. bilineata* last for 11 weeks (Grimm et al. 2000; Stacey et al. 2006; Raybould et al. 2007; Raybould and Vlachos 2010). For comparison, tests that assess mortality as a response to treatment with chemical pesticides can be substantially shorter (e.g., 2 days in the case of adult honeybees, USEPA 1996c).

2. *The selected surrogate and its life-stages.* The duration of the experiment depends on the test organism's biology and in particular its rate of development under the specific experimental conditions. For example, developmental time until adulthood is shorter for *Orius* species (Stacey et al. 2006; Duan et al. 2007) than for *Poecilus* species (Duan et al. 2006; Stacey et al. 2006).

3. *Suitability of the test system.* To ensure the reliability of the obtained results, test organisms should not show unacceptably high mortalities in the negative controls. This ensures that potential background effects (including quality of the test organisms and the suitability of the test system and environmental test conditions) were negligible and did not affect the observed treatment effects. Principles of basic toxicity testing dictate that test organisms should be healthy and of high quality and not otherwise stressed by factors other than the stressor in question (i.e., the arthropod-active protein) (Klaasen 2001). Therefore, negative control thresholds (e.g., maximum mortality or minimum fecundity) are generally defined for a specific organism and test protocol above which test results should be discarded (see also below section on "Control substances"). In general, a study should be terminated if control mortality exceeds a pre-defined threshold. That may mean that data are not used, or that the only data used are those prior to the control

mortality being exceeded. Whether to use those data will depend on the guideline being followed and when the control mortality criterion was exceeded. For example, if the criterion is exceeded after 21 days, it may be acceptable to draw reliable conclusions from the data for day 20 (e.g., if one is working with a *Bt* protein that is known to have effects on susceptible insects after 5 days); on the other hand, if the criterion is exceeded after 3 days, the study should be started again.

4. *Characteristics of the test substance.* Knowledge available on the impact of the test substance on a range of sensitive target organisms should be considered to demonstrate the spectrum of activity. For example, it has been recommended that test durations for *Bt* Cry proteins should be a minimum of 5 days, but preferably 7–14 days, in light of the time period for the proteins to demonstrate toxicity against some target pests (Rose 2007).

One should note that the period during which the test organism is actually exposed to the test substance can be shorter than the observation period (i.e., the duration of the experiment). One example is the honey bee larval test. In the bee test, young larvae are dosed with test substance in their brood cell only once at the start of the experiment, and adult emergence is measured after about 18 days (see references in OECD 2007). Another example is the protocol established to assess fecundity of *A. bilineata* (Grimm et al. 2000; Stacey et al. 2006) where adult beetles are treated with the test substance and provided with fly pupae to lay their eggs (the larvae develop parasitically within the fly pupae). After removal of the treated beetles, the fly pupae are monitored for a further 6 weeks to record the number of emerging second-generation beetles.

Control substances

Negative controls

The reason for using negative controls is to assess the natural background effects on the measurement endpoints within the test system. The inclusion of negative controls allows an assessment to be made of how the test system and test conditions, including the carrier, are influencing the mortality, development, and/or

behavior of the non-target arthropod tested. Thus, negative controls assist in determining whether the observed effects are related or not to the treatment.

When choosing the appropriate negative control treatment, it is important to consider the potential effects of the diet (or any other carrier) in which the test substance is delivered. Sometimes several negative control diets may be required. For example, in simple test diets, such as sugar solutions to which high concentrations of a protein test substance have been added, the inherent nutritional value of the test substance can affect the test results. In such cases, it may be appropriate to not only include an untreated control diet treatment but also to amend the negative control diet with an inert or heat-deactivated supplement such as bovine serum albumin to ensure nutritional equivalence (Brødsgaard et al. 2003; Bell et al. 2004).

The choice of the appropriate negative control is particularly critical when GE plant material is used to deliver the test substance. For GE plant material, typically material of the unmodified (non-transformed) near-isoline is used as the negative control in order to rule out effects of variability in plant composition on defined measurement endpoints. Genetic variation across plant varieties can cause differences in a range of plant compounds (e.g., Ridley et al. 2002; Zurbrügg et al. 2010), that may affect non-target organisms (e.g., Escher et al. 2000; Wandeler et al. 2002; Jensen et al. 2010; Knecht and Nentwig 2010). These differences are part of the normal variation in a crop. Consequently, it may be necessary to add several negative control reference lines to establish the normal response variation of the test NTA in the crop. This allows the assessor to set any observed effects into context.

The mortality observed in the negative control treatment group is a strong indicator of whether or not an appropriate study design has been used. Acceptable control mortalities need to be defined for any specific test as has been done for the standardized protocols for the acute testing of plant protection products (e.g., Candolfi et al. 2000). For example USEPA guidelines (USEPA 1996b) suggest terminating the test when control mortalities rise above 20% (Rose 2007). It has been recognized that higher control mortalities are expected and might be acceptable where assessment of arthropod-active proteins is required across multiple life-stages of

the test organism, thus requiring longer test durations (Romeis et al. 2008). In any case, Abbott's correction (Abbott 1925) should be applied to correct the treatment results for the mortality observed in the negative control group, and both corrected and observed mortality should be reported.

Positive controls

Positive control compounds are particularly useful for test protocol development and standardization (e.g., Duan et al. 2007). They may also be required for regulatory early-tier studies since they fulfill specific roles (see below). Consequently, the selection of appropriate positive controls requires careful consideration. In general, positive controls and test substances should have similar properties in terms of their route of toxicity (e.g., oral versus dermal) and behavior-modifying properties (e.g., repellent or anti-feedant properties).

There are several purposes of positive control treatments:

1. *Determine whether or not the test substance was actually ingested.* For this purpose any orally-active arthropod toxin can be used as a control. These control substances need to be provided to the test organisms in a similar way (i.e., in an artificial diet) as the test substance. Effects observed in the measurement endpoints indicate that the control substance has been ingested and thus provide indirect evidence for the ingestion of the test substance. Examples are stomach poisons such as the growth regulator teflubenzuron (Stacey et al. 2006), potassium arsenate (Duan et al. 2006, 2007), or the proteinase inhibitor E-64 (Duan et al. 2007). Alternatively, ingestion of test substance could also be confirmed by immuno-assays (e.g., ELISA test) of the treated arthropods (Vojtech et al. 2005; Li et al. 2008; Chen et al. 2009; Meissle and Romeis 2009b; Li and Romeis 2010; Álvarez-Alfageme et al. 2010) or their frass (Brandt et al. 2004; Christeller et al. 2005; Mulligan et al. 2010), by incorporation of a dye into the prepared diet and subsequent examination of the diet uptake (Rodrigo-Simón et al. 2006), by labeling the protein of interest with a fluorescent compound such as rhodamine and confirm its uptake by the test organisms (Hogervorst et al. 2006), or by simply determining the weight of the test arthropods prior to and after exposure to the test substance (Romeis et al. 2004; Li et al. 2008).

2. *Proof that the test system works (demonstrate that the test system is able to detect treatment effects).* Positive control compounds can be selected and their concentrations adjusted to show that the defined effect sizes are detectable within the experimental set-up. For example, snowdrop lectin has been used to determine whether sublethal effects on development time and fecundity can be detected (Lawo and Romeis 2008; Li et al. 2008; Álvarez-Alfageme et al. 2010). Such carefully selected positive controls may thus replace statistical power analyses and effect size calculations since they provide evidence that the effect of interest would have been detectable.

3. *Allow comparison to other test results.* Positive controls may function as useful references to permit comparison of experiments that have been conducted previously (e.g., to establish the sensitivity of the NTA and to establish validity of the assay), or across multiple laboratories.

Statistical considerations

Consultation with a skilled statistician conversant in environmental toxicology testing before an experiment is conducted should eliminate most common design problems. Appropriate statistical methods and statistical power, i.e., the probability of finding a difference that does exist, must be employed to reach meaningful conclusions.

Ideally, sample size calculations should be completed prior to the start of the experiment. This should be done to ensure that the assay is sufficiently replicated to detect a pre-defined effect size (e.g., 50% used by US EPA, Rose 2007) with appropriate statistical power (e.g., Duan et al. 2006, 2008a). A level of 80% power at an alpha level of 0.05 is usually considered acceptable (Rose 2007; Perry et al. 2009). Alternatively, retrospective power analyses may be conducted on non-significant results after a study has been completed (Steidl et al. 1997; Thomas 1997; Hoffmeister et al. 2006). For example, using the recorded mean and standard deviation of the negative control treatment and the true sample sizes tested in a particular study, one can calculate (a) the difference between the treatment and negative control that would have been detectable (given that $\alpha = 0.05$ and a power of 80%), or (b) the power achieved for a defined detectable difference (e.g., a 20% effect) (e.g., Marvier 2002; Lundgren and Wiedenmann 2002; Li et al. 2008; Meissle and Romeis 2009b).

The following statistical approaches are commonly used in laboratory non-target testing. For MHD tests, analysis of variance (ANOVA) and proportion tests (z test) are suitable (Candolfi et al. 2000; Rose 2007). If a threshold of activity is used as the criterion to trigger higher-tier tests (e.g., ≥50% mortality), then proportion tests can be used to directly evaluate if the experimental result observed when the NTA is exposed to a limit dose is significantly lower than this threshold. Since the null hypothesis, in this case, is that the experimental result is greater than this threshold (assuming control mortality criteria are met), rejection of the null hypothesis at a given alpha level, typically 0.05, provides 95% power in the conclusion that a result is less than the threshold (Rose 2007). One-sided tests may be appropriate for such considerations. For proportion testing, use of an alternative null hypothesis may also be acceptable. For cases where a threshold is not established, e.g., growth or reproduction endpoints or studies using plant material, use of ANOVA to compare treatment and the relevant comparator (control) may be most appropriate (Chapman et al. 1996). For dose-response studies, probit analysis, generalized probit analysis, logistic regression, and moving average-angle methods are suitable (Grimm et al. 2001). The robustness of the results should be documented by providing the 95% confidence or fiducial limits and the statistical significance of the fit of the data to the regression model.

See also for additional statistical advice and for guidance concerning the choice of testing procedures: guidance documents published by EPPO (EPPO 2007), the Society of Environmental Toxicology and Chemistry (SETAC) (Chapman et al. 1996), and the US Environmental Protection Agency (USEPA) (Rose 2007).

Conclusions

A sound environmental risk assessment is essential to evaluate the likelihood and seriousness of harm to NTAs that may result from cultivation of a GE arthropod-resistant crop. In such assessments, it is necessary to test for the potential of the arthropod-

active protein to have adverse effects on NTAs. Effective assessment of adverse effects follows a tiered approach that starts with laboratory studies under worst-case exposure conditions; such studies have a high ability to detect adverse effects on non-target species. As not all NTAs can be tested, early-tier laboratory studies should accurately determine the effects on surrogate non-target arthropods (selected depending on the scope of the risk assessment) of known concentrations of the test substance. In most cases, the test substance will be a purified protein produced in microbial expression systems, or, alternatively, GE plant tissue.

Good study design seeks to minimize the probability of erroneous results: false negatives—the failure to detect adverse effects of substances that are potentially harmful in the field, and false positives—the detection of adverse effects when the substance is unlikely to be harmful in the field. Thus, reliable test systems should adhere to relevant test protocol design criteria to avoid erroneous results (Box 1). Such erroneous results may arise if the conduct of the test introduces bias, or exposes the test NTAs to conditions that are significantly different from those under which the test is known to be reliable. Because some regulatory jurisdictions require that early-tier NTA studies be conducted under GLP, scientists conducting NTA studies as part of regulatory submissions should first determine if a GLP requirement exists.

Confidence in a conclusion of no adverse effect on a species (i.e., the avoidance of false negatives), and confidence in extrapolating that conclusion to other species, depends upon the ability of the study to detect such effects. Adhering to the principles and recommendations outlined in this paper should increase confidence in the results of early-tier laboratory studies, and thereby reduce data requirements for stressors that pose low risk. If adverse effects are detected in such studies, the results should be easier to interpret and higher-tier studies for GE crops producing those substances can be designed.

The recommendations and associated guidance elaborated in this document thus provide a sound scientific foundation for experimenters conducting early-tier NTA tests. These will also facilitate the reproduction of a study, peer review of such tests by others in the scientific community, and will benefit

Box 1 Study design criteria for NTA laboratory studies

(a) The test substance must be well characterized and described. This includes the source and purity of the arthropod-active protein, and its stability and homogeneity in the carrier through which it is provided to the test organism

(b) The test substances must be biochemically and functionally equivalent to the protein or other active ingredient produced in the GE crop

(c) The bioactivity of the test substances, as provided to the test organisms, must be established (e.g., in sensitive insect bioassays)

(d) Test organisms should be exposed to high concentrations of the test substance relative to predicted exposures in the field (if possible) or dose-response studies should be performed

(e) Exposure of the test organisms to the test substance should be confirmed by, for example, use of a positive control and diet analysis to measure the concentration of the test substance

(f) Endpoints should be measured that are likely to indicate the possibility of adverse effects on the abundance of NTAs or other assessment endpoints. Risk assessors should agree on how to interpret and use these data in the risk assessment. Determination of the measurement endpoint(s) should consider the knowledge about the impact of the arthropod-active protein on the target organisms, knowledge about the biology of the selected NTA species and life-stages, and the availability of reliable test protocols

(g) The number of replicates in the study should be such that defined effect sizes can be detected with sufficient statistical power

(h) Negative control treatments must be included to assess the suitability of the test system, the organisms (e.g., health) and the test conditions, and to evaluate potential effects of the matrix or formulation in which the test substance is delivered. Test results from assays with unacceptable high negative control mortality should be discarded

(i) Positive control treatments should be included, where feasible, to demonstrate that the test system is able to detect treatment effects

regulatory authorities by enhancing the quality of information generated for use in risk assessments.

Annex 1. Glossary of terms

Arthropod-active protein:	The active protein in a GE plant that causes the intended effects on the target organisms (i.e., the arthropod pests).
Assessment endpoint:	An explicit expression of the environmental value that is to be protected. Operationally, it is defined by an identified environmental entity of value (e.g., beneficial arthropods) that is susceptible to harm and an attribute that provides evidence of harm (e.g., arthropod abundance).
Bacillus thuringiensis (*Bt*):	A common soil bacterium, notable for its ability to produce proteins which are toxic to certain insect groups (see Cry proteins).
Carrier:	Matrix used to expose test organisms to a test substance. Carriers can be water, artificial diets, or arthropods containing the test substance. (see Diet).
Concentration:	Quantity of a test substance contained in a diet or other matrix (e.g., plant material), for example expressed as ng/mg, mg/kg, ng/ml, mg/L, etc.
Cry proteins:	A class of crystalline proteins produced by *Bacillus thuringiensis* bacteria that are toxic to specific insect taxa.
Diet:	In the context of laboratory-based toxicity testing of non-target organisms, material into which test substance is incorporated to enable oral exposure of test organisms.
Dose:	Total amount of a test substance administered to, taken, or absorbed by an organism.
EC_{50}:	Effective concentration; concentration of a test substance in a diet or any other matrix that has a defined effect (for example, growth inhibition) on fifty percent of the population of a test organism.
ED_{50}:	Effective dose; amount of a test substance that has a defined effect on fifty percent of the population of a test organism.
Environmental risk assessment (ERA):	The process of identifying significant risks to the environment, estimating the level of risk, and determining those risks that require measures to reduce the level of risk (USEPA 1998; Wolt et al. 2010).
Estimated Environmental Concentration (EEC):	Calculated concentrations of a substance in various environmental compartments based on calculations using maximum-exposure scenarios (IUPAC 2010).
Exposure:	The concentration or dose of an active protein (or any other stressor) encountered by an organism in the environment.
LC_{50}:	Lethal concentration; concentration of a test substance in a diet or any other matrix that causes the death of fifty percent of the population of a test organism.
LD_{50}:	Lethal dose; amount of a test substance which causes the death of fifty percent of the population of a test organism.
Lowest Observed Adverse Effect Concentration (LOAEC):	The lowest concentration of a test substance found by experiment or observation that causes an adverse effect on an organism.
Maximum Hazard Dose (MHD):	The dose chosen to represent an extreme exposure scenario, calculated using the EEC and incorporating an additional margin of exposure.
Measurement endpoint:	A quantifiable response to the changed characteristic of the plant (i.e., the stressor) that is related to the assessment endpoint. Examples include: mortality, growth (weight), development.
Mode of action (MOA):	Specific biochemical interaction or mechanism through which a compound produces its toxic effect.
Negative control:	Test organisms put through an experimental protocol that are not exposed to the test substance, but may be exposed to the carrier (e.g., an artificial diet) in which the test substance is contained. This is a test of the protocol that helps minimize false positives.
No Observed Adverse Effects Concentration (NOAEC):	The highest concentration or dose used in a test at which no adverse effects were observed.

Non-target arthropod (NTA):	An arthropod species that is not intended to be affected by the trait under consideration. Agronomically, a non-target organism is any organism associated with the crop that does not cause economical damage. Note that some scientists differentiate non-target pests (plant pests not targeted for control by the trait) from non-target organisms or arthropods, which are generally considered beneficial.
Positive control:	Test organisms put through an experimental protocol that are exposed to a substance known to have an effect (usually toxicity). This is used to show that the test organisms are exposed to the test substance during the test and to minimize false negatives.
Problem formulation:	The first phase of environmental risk assessment which includes a preliminary description of exposure and environmental effects, scientific data and data needs, key factors to be considered, and the scope and objectives of the assessment. This phase produces the risk hypotheses, conceptual model and analysis plan, around which the rest of the assessment develops (Raybould 2006; Wolt et al. 2010).
Protection goals:	The objectives of environmental policies, typically defined in law or regulations.
Reconstructability:	A property of a study so that other scientists or regulators can completely reassess the experiment at a later date. This is achieved by recording and achieving the protocols, data and execution of a study. Study reconstructability is a requirement under GLP.
Risk assessment:	The scientific determination of the probability of specified harmful effects under a given set of conditions, using both effect characterization and exposure data.
Risk hypothesis:	A tentative explanation of how the proposed actions, such as the cultivation of GE crops, may cause harm.
Sensitive insect bioassays:	The use of a known sensitive organism to provide evidence that the test substance under consideration has retained its biological activity.
Stressor:	Any physical, chemical, or biological entity (e.g., an arthropod-active protein) that can induce an adverse response in a receptor (e.g., a NTA).
Surrogate species:	A species selected for laboratory testing because it represents a taxonomic or functional group of organisms that should be addressed in the risk assessment.
Target organism:	A species that is intended to be adversely affected by the trait under consideration.
Test substance:	A purified arthropod-active protein or a GE plant tissue (e.g., pollen) that contains that protein. In general, test substance is referred to as the material that the test organism is exposed to.
Tiered risk assessment:	Risk hypotheses are systematically addressed within a framework of tests that proceeds from conservative laboratory-based tests to more complex and more realistic semi-field and field studies, if required. The assessment is iterative and can stop as soon as sufficient information is compiled to corroborate the risk hypothesis with sufficient confidence. Movement through the framework is triggered by results obtained at earlier tiers.
Trait:	A characteristic of an organism that manifests itself in the phenotype (e.g., insect-resistance). Traits may be the result of a single gene or may be polygenic, resulting from the simultaneous expression of more than one gene.
Trophic level:	An organism's place within a food chain, defined by the organisms it feeds upon. Tritrophic studies, for example, refer to interactions among three different trophic levels, e.g., plant, herbivore and natural enemy (parasitoid or predator).

References

Abbott WS (1925) A method of computing the effectiveness of an insecticide. J Econ Entomol 18:265–267

Álvarez-Alfageme F, Ferry N, Castañera P, Ortego F, Gatehouse AMR (2008) Prey mediated effects of Bt maize on fitness and digestive physiology of the red spider mite predator *Stethorus punctillum* Weise (Coleoptera: Coccinellidae). Transgenic Res 17:943–954

Álvarez-Alfageme F, Bigler F, Romeis J (2010) Laboratory toxicity studies demonstrate no adverse effects of Cry1Ab and Cry3Bb1 to larvae of *Adalia bipunctata* (Coleoptera: Coccinellidae): the importance of study design. Transgenic Res. doi:10.1007/s11248-010-9430-5

Aupinel P, Fortini D, Michaud B, Marolleau F, Tasei J-N, Odoux J-F (2007) Toxicity of dimethoate and fenoxycarb to honey bee brood (*Apis mellifera*), using a new in vitro standardized feeding method. Pest Manag Sci 63:1090–1094

Aupinel P, Fortini D, Michaud B, Medrzycki P, Padovani E, Przygoda D, Maus C, Charriere J-D, Kilchenmann V, Riessberger-Galle U, Vollmann JJ, Jeker L, Janke M, Odoux J-F, Tasei J-N (2009) Honey bee brood ring-test: method for testing pesticide toxicity on honeybee brood in laboratory conditions. Julius-Kühn-Archiv 423:96–102

Babendreier D, Kalberer N, Romeis J, Fluri P, Bigler F (2004) Pollen consumption in honey bee larvae: a step forward in the risk assessment of transgenic plants. Apidologie 35:293–300

Bai YY, Jiang MX, Cheng JA (2005) Effects of transgenic cry1Ab rice pollen on fitness of *Propylea japonica* (Thunberg). J Pest Sci 78:123–128

Barrett KL, Grandy N, Harrison EG, Hassan S, Oomen P (eds) (1994) Guidance document on regulatory testing procedures for pesticides with non-target arthropods. From the workshop European standard characteristics of beneficials regulatory testing (ESCORT), Society of Environmental Toxicology and Chemistry Europe, Brussels

Bell HA, Down RE, Fitches EC, Edwards JP, Gatehouse AMR (2003) Impact of genetically modified potato expressing plant-derived insect resistance genes on the predatory bug *Podisus maculiventris* (Heteroptera: Pentatomidae). Biocontrol Sci Technol 13:729–741

Bell HA, Kirkbride-Smith AE, Marris GC, Edwards JP, Gatehouse AMR (2004) Oral toxicity and impact on fecundity of three insecticidal proteins on the gregarious ectoparasitoid *Eulophus pennicornis* (Hymenoptera: Eulophidae). Agric For Entomol 6:215–222

Brandt SL, Coudron TA, Habibi J, Brown GR, Ilagan OM, Wagner RM, Wright MK, Backus EA, Huesing JE (2004) Interaction of two *Bacillus thuringiensis* δ-endotoxins with the digestive system of *Lygus hesperus*. Curr Microbiol 48: 1–9

Brødsgaard HF, Brødsgaard CJ, Hansen H, Lövei G (2003) Environmental risk assessment of transgene products using honey bee (*Apis mellifera*) larvae. Apidologie 34: 139–145

CAC (2003) Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants. CAC/GL 45-2003. Codex Alimentarius Commission (CAC), Rome. http//www.who.int/foodsafety/biotech/en/codex_guidelines_plants.pdf. Accessed 9 Sep 2010

Candolfi MP, Blümel S, Forster R, Bakker FM, Grimm C, Hassan SA, Heimbach U, Mead-Briggs MA, Reber B, Schmuck R, Vogt H (eds) (2000) Guidelines to evaluate side-effects of plant protection products to non-target arthropods. International Organization for Biological and Integrated Control of Noxious Animals and Weeds, West Palearctic Regional Sectionnnnn (IOBC/WPRS), Gent

Candolfi MP, Barrett KL, Campbell PJ, Forster R, Grandy N, Huet M-C, Lewis G, Oomen PA, Schmuck R, Vogt H (eds) (2001) Guidance document on regulatory testing and risk assessment procedures for plant protection products with non-target arthropods. From the workshop European standard characteristics of non-target arthropod regulatory testing (ESCORT 2), Society of Environmental Toxicology and Chemistry Europe, Brussels

Caro TM, O'Doherty G (1999) On the use of surrogate species in conservation biology. Conserv Biol 13:805–814

Carstens KL, Hayter K, Layton RJ (2010) A perspective on problem formulation and exposure assessment of transgenic crops. IOBC/WPRS Bull 52:23–30

CDPR (2009) A standardized procedure for the in vitro rearing of honey bee larvae (developed by Zachary Huang). California Department of Pesticide Regulation, http://www.cdpr.ca.gov/docs/registration/reevaluation/chemicals/neonicotinoids.htm. Accessed 9 Sep 2010

Chapman PF, Crane M, Wiles J, Noppert F, McIndoe E (1996) Improving the quality of statistics in regulatory ecotoxicity tests. Ecotoxicology 5:169–186

Chen M, Zhao J-Z, Collins HL, Earle ED, Cao J, Shelton AM (2008) A critical assessment of the effects of Bt transgenic plants on parasitoids. PLoS ONE 3(5):e2284

Chen M, Ye G-Y, Liu Z-C, Fang Q, Hu C, Peng Y-F, Shelton AM (2009) Analysis of Cry1Ab toxin bioaccumulation in a food chain of Bt rice, an herbivore and a predator. Ecotoxicology 18:230–238

Christeller JT, Malone LA, Todd JH, Marshall RM, Burgess EPJ, Philip BA (2005) Distribution and residual activity of two insecticidal proteins, avidin and aprotinin, expressed in transgenic tobacco plants, in the bodies and frass of *Spodoptera litura* larvae following feeding. J Insect Physiol 51:1117–1126

Cohen AC (2004) Insect diets: science and technology. CRC Press, Boca Raton

Dively GP, Rose R, Sears MK, Hellmich RL, Stanley-Horn DE, Calvin DD, Russoe JM, Anderson PL (2004) Effects on monarch butterfly larvae (Lepidoptera: Danaidae) after continuous exposure to Cry1Ab-expresing corn during anthesis. Environ Entomol 33:1116–1125

Duan JJ, Head G, McKee MJ, Nickson TE, Martin JW, Sayegh FS (2002) Evaluation of dietary effects of transgenic corn pollen expressing Cry3Bb1 protein on a non-target ladybird beetle, *Coleomegilla maculata*. Ent Exp Appl 104: 271–280

Duan JJ, Paradise MS, Lundgren JG, Bookout JT, Jian C, Wiedenmann RN (2006) Assessing non-target impacts of Bt corn resistant to corn rootworms: Tier-1 testing with larvae of *Poecilus chalcites* (Coleoptera: Carabidae). Environ Entomol 35:135–142

Duan JJ, Huesing J, Teixeira D (2007) Development of Tier-I toxicity assays for *Orius insidiosus* (Heteroptera: Anthocoridae) for assessing the risk of plant-incorporated protectants to nontarget heteropterans. Environ Entomol 36:982–988

Duan JJ, Teixeira D, Huesing JE, Jiang C (2008a) Assessing the risk to nontarget organisms from Bt corn resistant to corn rootworms (Coleoptera: Chrysomelidae): Tier-I testing with *Orius insidiosus* (Heteroptera: Anthocoridae). Environ Entomol 37:838–844

Duan JJ, Marvier M, Huesing J, Dively G, Huang ZY (2008b) A meta-analysis of effects of Bt crops on honey bees (Hymenoptera: Apidae). PLoS ONE 3(1):e1415

Duan JJ, Lundgren JG, Naranjo S, Marvier M (2010) Extrapolating non-target risk of Bt crops from laboratory to field. Biol Lett 6:74–77

Dutton A, Klein H, Romeis J, Bigler F (2002) Uptake of Bt-toxin by herbivores feeding on transgenic maize and consequences for the predator *Chrysoperla carnea*. Ecol Entomol 27:441–447

EFSA (2008) Guidance document for the risk assessment of genetically modified plants and derived food and feed by the Scientific Panel on Genetically Modified Organisms (GMO)—including draft document updated in 2008. European Food Safety Authority (EFSA), Parma. http://www.efsa.europa.eu/en/scdocs/scdoc/99.htm. Accessed 9 Sep 2010

EPPO (2007) EPPO guideline for the efficacy evaluation of plant protection products: Dsign and analysis of efficacy evaluation trials, PP 1/152(3). European Plant Protection Organisation, Paris

Escher N, Käch B, Nentwig W (2000) Decomposition of transgenic *Bacillus thuringiensis* maize by micro-organisms and woodlice *Porcellio scaber* (Crustacea: Isopoda). Basic Appl Ecol 1:161–169

Ferry N, Mulligan EA, Stewart CN, Tabashnik BE, Port GR, Gatehouse AMR (2006) Prey mediated effects of transgenic canola on beneficial, non-target, carabid beetle. Transgenic Res 15:501–514

Gao Y, Schafer BW, Collins RA, Herman RA, Xu X, Gilbert JR, Ni W, Langer VL, Tagliani LA (2004) Characterization of Cry34Ab1 and Cry35Ab1 insecticidal crystal proteins expressed in transgenic corn plants and *Pseudomonas fluorescens*. J Agric Food Chem 52:8057–8065

Gao Y, Fencil KJ, Xu X, Schwedler DA, Gilbert JR, Herman RA (2006) Purification and characterization of a chimeric Cry1F δ-endotoxin expressed in transgenic cotton plants. J Agric Food Chem 54:829–835

Garcia-Alonso M, Jacobs E, Raybould A, Nickson TE, Sowig P, Willekens H, van der Kouwe P, Layton R, Amijee F, Fuentes AM, Tencalla F (2006) A tiered system for assessing the risk of genetically modified plants to non-target organisms. Environ Biosafety Res 5:57–65

Glare TR, O'Callaghan M (2000) *Bacillus thuringiensis*: Biology, ecology and safety. Wiley, Chichester

Grimm C, Reber B, Candolfi MP, Drexler A, Maus C, Moreth L, Ufer A, Waltersdorfer A (2000) A test for evaluating the chronic effects of plant protection products on the rove beetle *Aleochara bilineata* Gyll. (Coleoptera: Staphylinidae) under laboratory and extended laboratory conditions. In: Candolfi MP et al (eds) Guidelines to evaluate side-effects of plant protection products to non-target arthropods. IOBC/WPRS, Gent, pp 1–12

Grimm C, Schmidli H, Bakker F, Browm K, Campbell P, Candolfi M, Chapmann P, Harrison EG, Mead-Briggs M, Schmuck R, Ufer A (2001) Use of standard toxicity tests with *Typhlodromus pyri* and *Aphidius rhopalosiphi* to establish a dose-response relationship. J Pest Sci 74:72–84

Grimm C, Reber B, Gray A, Candolfi MP (2002) Adaptation of standard regulatory non-target arthropod test methods for testing of plant protection products with special modes of action. IOBC/WPRS Bull 25(11):7–17

Harwood JD, Wallin WG, Obrycki JJ (2005) Uptake of Bt endotoxins by nontarget herbivores and higher order arthropod predators: molecular evidence from a transgenic corn agroecosystem. Mol Ecol 14:2815–2823

Head G, Brown CR, Groth ME, Duan JJ (2001) Cry1Ab protein levels in phytophagous insects feeding on transgenic corn: implications for secondary exposure risk assessment. Ent Exp Appl 99:37–45

Heckmann LH, Griffiths BS, Caul S, Thompson J, Pusztai-Carey M, Moar WJ, Andersen MN, Krogh PH (2006) Consequences for *Protaphorura armata* (Collembola: Onychiuridae) following exposure to genetically modified *Bacillus thuringiensis* (Bt) maize and non-Bt maize. Environ Pollut 142:212–216

Heimbach U, Wehling A, Barrett KL, Candolfi MP, Jäckel B, Kennedy PJ, Mead-Briggs M, Nienstedt KM, Römbke J, Schmitzer S, Schmuck R, Ufer A, Wilhelmy H (2000a) A method for testing effects of plant protection products on spiders of the genus *Pardosa* (Araneaa, Lycosidae) under laboratory conditions. In: Candolfi MP et al (eds) Guidelines to evaluate side-effects of plant protection products to non-target arthropods. IOBC/WPRS, Gent, pp 71–86

Heimbach U, Dohmen P, Barrett KL, Brown K, Kennedy PJ, Kleiner R, Römbke J, Schmitzer S, Schmuck R, Ufer A, Wilhelmy H (2000b) A method for testing effects of plant protection products on the carabid beetle *Poecilus cupreus* (Coleoptera: Carabidae) under laboratory and semi-field conditions. In: Candolfi MP et al (eds) Guidelines to evaluate side-effects of plant protection products to non-target arthropods. IOBC/WPRS, Gent, pp 87–106

Hellmich RL, Siegfried BD, Sears MK, Stanley-Horn DE, Mattila HR, Spencer T, Bidne KG, Daniels MJ, Lewis LC (2001) Monarch larvae sensitivity to *Bacillus thurningiensis*-purified proteins and pollen. Proc Nat Acad Sci USA 98:11925–11930

Hill RA, Sendashonga C (2003) General principles for risk assessment of living modified organisms: lessons from chemical risk assessment. Environ Biosafety Res 2:81–88

Hoffmeister TS, Babendreier D, Wajnberg E (2006) Statistical tools to improve the quality of experiments and data analysis. In: Bigler F, Babendreier D, Kuhlmann U (eds) Environmental impact of invertebrates for biological control of arthropods: methods and risk assessment. CABI Publishing, Wallingford, pp 222–240

Hogervorst PAM, Ferry N, Gatehouse AMR, Wäckers FL, Romeis J (2006) Direct effects of snowdrop lectin (GNA) on larvae of three aphid predators and fate of GNA after ingestion. J Insect Phys 52:614–624

Hogervorst PAM, Wäckers FL, Woodring J, Romeis J (2009) Snowdrop lectin (*Galanthus nivalis* agglutinin) in aphid honeydew negatively affects survival of a honeydew-consuming parasitoid. Agric For Entomol 11:161–173

IUPAC (2010) Glossary of terms used in toxicology. United States National Library of Medicine, Environmental Health and Toxicology Specialized Information Service. http://sis.nlm.nih.gov/enviro/iupacglossary/glossarye.html. Accessed 11 June 2010

Jensen PD, Dively GP, Swan CM, Lamp WO (2010) Exposure and nontarget effects of transgenic *Bt* corn debris in streams. Environ Entomol 39:707–714

Jesse LCH, Obrycki JJ (2002) Assessment of the non-target effects of transgenic Bt corn pollen and anthers on the milkweed tiger moth *Euchatias egle* Drury (Lepidoptera: Arctiidae). J Kansas Ent Soc 75:55–58

Klaasen CD (ed) (2001) Casarett and Doull's toxicology: the basic science of poisons, 6th edn. McGraw-Hill, New York

Knecht S, Nentwig W (2010) Effect of *Bt* maize on the reproduction and development of saprophagous Diptera over multiple generations. Basic Appl Ecol 11:346–353

Konrad R, Ferry N, Gatehouse AMR, Babendreier D (2008) Potential effects of oilseed rape expressing oryzacystatin-1 (OC-1) and of purified insecticidal proteins on larvae of the solitary bee *Osmia bicornis*. PLoS ONE 3(7):e2664

Lawo NC, Romeis J (2008) Assessing the utilization of a carbohydrate food source and the impact of insecticidal proteins on larvae of the green lacewing, *Chrysoperla carnea*. Biol Control 44:389–398

Lawo NC, Wäckers FL, Romeis J (2010) Characterizing indirect prey-quality mediated effects of a *Bt* crop on

predatory larvae of the green lacewing, *Chrysoperla carnea*. J Insect Phys 56:1702–1710

Li Y, Romeis J (2010) *Bt* maize expressing Cry3Bb1 does not harm the spider mite, *Tetranychus urticae*, or its ladybird beetle predator, *Stethorus punctillum*. Biol Control 53:337–344

Li W, Wu K, Wang X, Wang G, Guo Y (2005) Impact of pollen grains from Bt transgenic corn on the growth and development of Chinese tussah silkworm, *Antheraea pernyi* (Lepidoptera: Saturniidae). Environ Entomol 34:922–928

Li Y, Meissle M, Romeis J (2008) Consumption of Bt maize pollen expressing Cry1Ab or Cry3Bb1 does not harm adult green lacewings, *Chrysoperla carnea* (Neuroptera: Chrysopidae). PLoS ONE 3(8):e2909

Li Y, Meissle M, Romeis J (2010) Use of maize pollen by adult *Chrysoperla carnea* (Neuroptera: Chrysopidae) and fate of Cry proteins in Bt-transgenic varieties. J Insect Phys 56:157–164

Ludy C, Lang A (2006) Bt maize pollen exposure and impact on the garden spider, *Araneus diadematus*. Ent Exp Appl 118:145–156

Lundgren JG, Wiedenmann RN (2002) Coleopteran-specific Cry3Bb toxin from transgenic corn pollen does not affect the fitness of a nontarget species, *Coleomegilla maculata* DeGeer (Coleoptera: Coccinellidae). Environ Entomol 31:1213–1218

Lundgren JG, Wiedenmann RN (2004) Nutritional suitability of corn pollen for the predator *Coleomegilla maculata* (Coleoptera: Coccinellidae). J Insect Physiol 50:567–575

Lundgren JG, Huber A, Wiedenmann RN (2005) Quantification of consumption of corn pollen by the predator *Coleomegilla maculata* (Coleoptera: Coccinellidae) during anthesis in an Illinois cornfield. Agric For Entomol 7:53–60

Malone LA, Burgess EPJ, Stefanovic D (1999) Effects of a *Bacillus thuringiensis* toxin, two *Bacillus thuringiensis* biopesticide formulations, and a soybean trypsin inhibitor on honey bee (*Apis mellifera* L.) survival and food consumption. Apidologie 30:465–473

Marvier MA (2002) Improving risk assessment for nontarget safety of transgenic crops. Ecol Appl 12:1119–1124

Meissle M, Romeis J (2009a) Insecticidal activity of Cry3Bb1 expressed in Bt maize on larvae of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). Ent Exp Appl 131:308–319

Meissle M, Romeis J (2009b) The web-building spider *Theridion impressum* (Araneae: Theridiidae) is not adversely affected by Bt maize resistant to corn rootworms. Plant Biotech J 7:645–656

Mulligan EA, Ferry N, Jouanin L, Romeis J, Gatehouse AMR (2010) Characterisation of adult green lacewing (*Chrysoperla carnea*) digestive physiology: impact of a cysteine protease inhibitor and a synthetic pyrethroid. Pest Manag Sci 66:325–336

Naranjo SE (2009) Impacts of Bt crops on non-target organisms and insecticide use patterns. CAB Rev Persp Agric Vet Sci Nutrit Nat Resour 4(011)

Obrist LB, Dutton A, Albajes R, Bigler F (2006a) Exposure of arthropod predators to Cry1Ab toxin in Bt maize fields. Ecol Entomol 31:143–154

Obrist L, Dutton A, Romeis J, Bigler F (2006b) Biological activity of Cry1Ab toxin expressed by Bt maize following ingestion by herbivorous arthropods and exposure of the predator *Chrysoperla carnea*. Biocontrol 51:31–48

Obrist LB, Klein H, Dutton A, Bigler F (2006c) Assessing the effects of Bt maize on the predatory mite *Neoseiulus cucumeris*. Exp Appl Acar 38:125–139

OECD (1998) OECD series on principles of good laboratory practice and compliance monitoring. Organisation for Economic Co-operation and Development, Paris. http://www.oecd.org/document/63/0,3343,en_2649_34381_2346175_1_1_1_1,00.html. Accessed 9 Sep 2010

OECD (2005) Good laboratory practice. OECD principles and guidance for compliance monitoring. Organisation for Economic Co-operation and Development, Paris

OECD (2007) Consensus document on safety information on transgenic plants expressing Bacillus thuringiensis—derived insect control protein. Series on Harmonisation of Regulatory Oversight in Biotechnology, No. 42. Organisation for Economic Co-operation and Development, Paris. http://www.olis.oecd.org/olis/2007doc.nsf/LinkTo/NT00002DF6/$FILE/JT03230592.PDF. Accessed 9 Sep 2010

OGTR (2009) Risk analysis framework. Office of the Gene Technology Regulator, Canberra. http://www.health.gov.au/internet/ogtr/publishing.nsf/Content/riskassessments-1. Accessed 9 Sep 2010

Perry JN, ter Braak CJF, Dixon PM, Duan JJ, Hails RS, Huesken A, Lavielle M, Marvier M, Scardi M, Schmidt K, Tothmeresz B, Schaarschmidt F, van der Voet H (2009) Statistical aspects of environmental risk assessment of GM plants for effects on non-target organisms. Environ Biosafety Res 8:65–78

Raybould A (2006) Problem formulation and hypothesis testing for environmental risk assessments of genetically modified crops. Environ Biosafety Res 5:119–125

Raybould A (2007) Environmental risk assessment of genetically modified crops: general principles and risks to non-target organisms. BioAssay 2:8

Raybould A, Quemada H (2010) *Bt* crops and food security in developing countries: realised benefits, sustainable use and lowering barriers to adoption. Food Sec 2:247–259

Raybould A, Vlachos D (2010) Non-target organism effects tests on Vip3A and their application to the ecological risk assessment for cultivation of MIR162 maize. Transgenic Res. doi:10.1007/s11248-010-9442-1

Raybould A, Stacey D, Vlachos D, Graser G, Li X, Joseph R (2007) Non-target organisms risk assessment of MIR604 maize expressing mCry3A for control of corn rootworms. J Appl Ent 131:391–399

Raybould A, Tuttle A, Shore S, Stone T (2010) Environmental risk assessments for transgenic crops producing output trait enzymes. Transgenic Res 19:595–609

Ridley WP, Sidhu RS, Pyla PD, Nemeth MA, Breeze ML, Astwood JD (2002) Comparison of the nutritional profile of glyphosate-tolerant corn event NK603 with that of conventional corn (*Zea mays* L.). J Agric Food Chem 50:7235–7243

Rodrigo-Simón A, de Maagd RA, Avilla C, Bakker PL, Molthoff J, González-Zamora JE, Ferré J (2006) Lack of detrimental effects of *Bacillus thuringiensis* cry toxins on

the insect predator *Chrysoperla carnea*: a toxicological, histopathological, and biochemical analysis. Appl Environ Microbiol 72:1595–1603

Romeis J, Meissle M (2010) Non-target risk assessment of *Bt* crops—Cry protein uptake by aphids. J Appl Ent. doi: 10.1111/j.1439-0418.2010.01546.x

Romeis J, Babendreier D, Wäckers FL (2003a) Consumption of snowdrop lectin (*Galanthus nivalis* agglutinin) causes direct effects on adult parasitic wasps. Oecologia 134: 528–536

Romeis J, Battini M, Bigler F (2003b) Transgenic wheat with enhanced fungal resistance causes no effects on *Folsomia candida* (Collembola: Isotomidae). Pedobiologia 47:137–144

Romeis J, Dutton A, Bigler F (2004) *Bacillus thuringiensis* toxin (Cry1Ab) has no direct effect on larvae of the green lacewing *Chrysoperla carnea* (Stephens) (Neuroptera: Chrysopidae). J Insect Phys 50:175–183

Romeis J, Meissle M, Bigler F (2006) Transgenic crops expressing *Bacillus thuringiensis* toxins and biological control. Nat Biotechnol 24:63–71

Romeis J, Bartsch D, Bigler F, Candolfi MP, Gielkens MMC, Hartley SE, Hellmich RL, Huesing JE, Jepson PC, Layton R, Quemada H, Raybould A, Rose RI, Schiemann J, Sears MK, Shelton AM, Sweet J, Vaituzis Z, Wolt JD (2008) Assessment of risk of insect-resistant transgenic crops to nontarget arthropods. Nat Biotechnol 26:203–208

Romeis J, Meissle M, Raybould A, Hellmich RL (2009) Impact of insect-resistant transgenic crops on above-ground non-target arthropods. In: Ferry N, Gatehouse AMR (eds) Environmental impact of genetically modified crops. CAB International, Wallingford, pp 165–198

Rose RI (ed) (2007) White paper on tier-based testing for the effects of proteinaceous insecticidal plant-incorporated protectants on non-target invertebrates for regulatory risk assessment. USDA-APHIS and US Environmental Protection Agency, Washington, DC, USA. http://www.epa.gov/pesticides/biopesticides/pips/non-target-arthropods.pdf. Accessed 9 Sep 2010

Rose R, Dively GP, Pettis J (2007) Effects of Bt corn pollen on honey bees: emphasis on protocol development. Apidologie 38:368–377

Schuler TH, Potting RPJ, Denholm I, Clark SJ, Clark AJ, Stewart CN, Poppy GM (2003) Tritrophic choice experiments with Bt plants, the diamond moth (*Plutella xylostella*) and the parasitoid *Cotesia plutellae*. Transgenic Res 12:351–361

Schuler TH, Denholm I, Clark SJ, Stewart CN, Poppy GM (2004) Effects of Bt plants on the development and survival of the parasitoid *Cotesia plutellae* (Hymenoptera: Braconidae) in susceptible and Bt-resistant larvae of the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). J Insect Phys 50:435–443

Shade RE, Schroeder HE, Pueyo JJ, Tabe LM, Murdock LL, Higgins TJV, Chrispeels MJ (1994) Transgenic pea seeds expressing the α-amylase inhibitor of the common bean are resistant to Bruchid beetles. Nat Biotechnol 12:793–796

Shirai Y, Takahashi M (2005) Effects of transgenic Bt corn pollen on a non-target lycaenid butterfly, *Pseudozizeeria maha*. Appl Entomol Zool 40:151–159

Stacey D, Graser G, Mead-Briggs M, Raybould A (2006) Testing the impact on non-target organisms of insecticidal proteins expressed in transgenic crops. IOBC/WPRS Bull 29(5):171–179

Stanley-Horn DE, Dively GP, Hellmich RL, Sears MK, Rose R, Jesse LCH, Losey JE, Obrycki JJ, Lewis LC (2001) Assessing the impact of Cry1Ab-expressing corn pollen on monarch butterfly larvae in field studies. Proc Nat Acad Sci USA 98:11931–11936

Steidl RJ, Hayes JP, Schauber E (1997) Statistical power analysis in wildlife research. J Wildl Manag 61:270–279

Thomas L (1997) Retrospective power analysis. Conserv Biol 11:276–280

Torres JB, Ruberson JR (2008) Interactions of *Bacillus thuringiensis* Cry1Ac toxin in genetically engineered cotton with predatory heteropterans. Transgenic Res 17:345–354

USEPA (1996a) Microbial pesticide test guidelines: OPPTS 885.0001 Overview for microbial pest control agents. EPA 712-C-96-280, February 1996. United States Environmental Protection Agency, Washington. http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series885.htm. Accessed 9 Sep 2010

USEPA (1996b) Microbial pesticide test guidelines: OPPTS 885.4000 Background for nontarget organism testing of microbial pest control agents. EPA 712-C-96-328, February 1996. United States Environmental Protection Agency, Washington. http://www.epa.gov/ocspp/pubs/frs/publications/Test_Guidelines/series885.htm. Accessed 9 Sep 2010

USEPA (1996c) Ecological effects test guidelines. OPPTS 850.3020 Honey bee acute contact toxicity. EPA 712–C–96–147, April 1996. United States Environmental Protection Agency, Washington. http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/850_Ecological_Effects_Test_Guidelines/Drafts/850-3020.pdf. Accessed 9 Sep 2010

USEPA (1998) Guidelines for ecological risk assessment. EPA/630/R-95/002F, April 1998 Final. United States Environmental Protection Agency, Washington

USEPA (2000) Report of the FIFRA Scientific Advisory Panel Meeting, "Sets of scientific issues being considered by the Environmental Protection Agency regarding: section II—mammalian toxicity assessment guidelines for protein plant pesticides", June 6–7, 2000, Arlington. United States Environmental Protection Agency, Washington. http://www.epa.gov/scipoly/sap/meetings/2000/June/finbtmamtox.pdf. Accessed 9 Sep 2010

USEPA (2001) Biopesticide registration action document. *Bacillus thuringiensis* (Bt) plant-incorporated protectants. United States Environmental Protection Agency, Washington. http://www.epa.gov/oppbppd1/biopesticides/pips/bt_brad.htm. Accessed 9 Sep 2010

USEPA (2009) Title 40 Part 160 – Good Laboratory Practice Standards. United States Environmental Protection Agency, Washington. http://www.access.gpo.gov/nara/cfr/waisidx_07/40cfr160_07.html. Accessed 9 Sep 2010. http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr;sid=a86247d14007df5afe20b9989f6aa96f;rgn=div5;view=text;node=40%3A23.0.1.1.11;idno=40;cc=ecfr. Accessed 9 Sep 2010

Van Damme EJM, Allen AK, Peumans WJ (1987) Isolation and characterization of a lectin with exclusive specificity towards mannose from snowdrop (*Galanthus nivalis*) bulbs. FEBS Lett 215:140–144

Vojtech E, Meissle M, Poppy GM (2005) Effects of Bt maize on the herbivore *Spodoptera littoralis* (Lepidoptera: Noctuidae) and the parasitoid *Cotesia marginiventris* (Hymenoptera: Braconidae). Transgenic Res 14:133–144

Wandeler H, Bahylova J, Nentwig W (2002) Consumption of two *Bt* and six non-*Bt* corn varieties by the woodlouse *Porcellio scaber*. Basic Appl Ecol 3:357–365

WHO (2008) Uncertainty and data quality in exposure assessment: Part 2, Hallmarks of data quality in chemical exposure assessment. International Programme on Chemical Safety Harmonization Project Document No. 6. World Health Organisation, Geneva. http://www.who.int/ipcs/publications/methods/harmonization/exposure_assessment.pdf. Accessed 9 Sep 2010

Wolt JD, Keese P, Raybould A, Fitzpatrick JW, Burachik M, Gray A, Olin SS, Schiemann J, Sears M, Wu F (2010) Problem formulation in the environmental risk assessment for genetically modified plants. Transgenic Res 19:425–436

Wraight CL, Zangerl AR, Carroll MJ, Berenbaum MR (2000) Absence of toxicity of *Bacillus thuringiensis* pollen to black swallowtails under field conditions. Proc Natl Acad Sci USA 97:7700–7703

Yu L, Berry E, Croft BA (1997) Effects of *Bacillus thuringiensis* toxins in transgenic cotton and potato on *Folsomia candida* (Collembola: Isotomidae) and *Oppia nitens* (Acari: Orbatidae). J Econ Entomol 90:113–118

Zurbrügg C, Höhnemann L, Meissle M, Romeis J, Nentwig W (2010) Decomposition dynamics and structural plant components of genetically modified *Bt* maize leaves do not differ from leaves of conventional hybrids. Transgenic Res 19:257–267