

## FORUM

**Environmental Risk Assessment of Genetically Modified Crops:  
General Principles and Risks to Non-target Organisms**

ALAN RAYBOULD

*Syngenta, Jealott's Hill International Research Station, Bracknell, Berkshire RG42 6EY, UK.**E-mail: [alan.raybould@syngenta.com](mailto:alan.raybould@syngenta.com)*

---

*BioAssay 2:8 (2007)***Análise de Risco Ambiental de Plantas Geneticamente Modificadas: Princípios Gerais e Riscos a Organismos Não-Alvos**

**RESUMO** – O presente trabalho discute que a análise de risco deveria ser abordada segundo o modelo de desenvolvimento contínuo do saber científico proposto por Karl Popper. Nesse contexto, a análise de risco deveria começar com o problema e busca de respostas para esse problema mediante testes de hipóteses. A análise de risco sendo considerada como teste de hipóteses, a segurança não pode ser provada, porém pode ser indicada pelos testes de hipóteses que prevêm baixo risco. A confiabilidade na análise de risco é dada pelo rigor com que hipóteses de risco são testadas; sendo que os testes devem ser iniciados em condições mais prováveis para demonstrar que as hipóteses de risco são falsas. Se as hipóteses de risco são corroboradas nessas condições, há confiança de que os riscos impostos pelas plantas geneticamente modificadas são baixos. O aumento no rigor nos testes de hipóteses auxilia para justificar o estabelecimento da solicitação de dados adicionais, e pode reduzir o risco ambiental mediante prevenção de atrasos excessivos no registro de produtos ambientalmente benéficos.

**PALAVRAS-CHAVE** – Método científico, formulação de problema, teste de hipóteses.

**ABSTRACT** – This paper argues that risk assessment should be viewed as conforming to the model of the continuous development of scientific knowledge proposed by Karl Popper. As such, a risk assessment should begin with a problem and search for answers to that problem by testing hypotheses. Regarding a risk assessment as hypothesis testing recognises that safety cannot be proved, but can be indicated by tests of hypotheses that predict low risk. Confidence in the risk assessment is provided by the rigour with which the risk hypotheses are tested; it follows that testing should begin under conditions most likely to reveal that the risk hypothesis is false. If the risk hypothesis is corroborated under those conditions, there can be confidence that the risks posed by the genetically modified plant are low. Application of a criterion of increased rigour for hypothesis testing helps to establish whether requests for additional data are justified, and may reduce environmental risk by preventing undue delay in the registration of environmentally beneficial products.

**KEYWORDS** – Scientific method, problem formulation, hypothesis testing.

---

The cultivation of genetically modified (GM) crops is strictly regulated worldwide. Before GM seeds can be sold and cultivated without restriction, a permit or licence must be obtained from a regulatory authority. The decision to license a GM crop for commercial cultivation is based on risk analysis, which judges whether the risk from use of the GM crop is acceptable. Risk analysis comprises two activities: risk assessment, a determination of the probability of specified harmful effects; and decision-making, the evaluation of whether the risk, and the uncertainty associated with its estimation, is acceptable. Acceptability depends on the objectives of public policy, along with the ability to

manage and communicate the risk (Wolt & Peterson 2000; Johnson *et al.* 2007).

Risk assessment cannot be separated completely from the other aspects of risk analysis because policy should determine which effects are considered in the risk assessment (Stern & Fineberg 1996), and because the risk assessment should seek to inform policy, not necessarily to increase general scientific knowledge (Hill and Sendashonga 2003; Raybould 2006). Nevertheless, most students of risk analysis consider risk assessment to be scientific, and as such it should follow the scientific method (*e.g.*, Power & Adams 1997; Patton 1998; Wolt & Petersen 2000; Johnson *et*

*al.* 2007). This paper suggests some general principles for design of risk assessments and illustrates the application of those principles with a conceptual model for assessing the risks of GM crops to non-target organisms<sup>1</sup>. A companion paper (Raybould 2007) applies these principles to a conceptual model for assessing the risks of gene flow from GM crops to wild relatives.

## The Structure of Environmental Risk Assessments

### Production of Scientific Knowledge

Environmental risk assessments for any proposed action follow the same structure and are simple in concept: decide what needs protection; assess how the proposed action may cause harm to the entities requiring protection; and collect data to predict the probability and magnitude of harm following that action. Once the prediction is made, it may be decided that the probability of harmful effects is known with sufficient certainty to allow a decision on whether to permit the proposed action. If there is insufficient certainty, further data may be collected to improve the characterization of risk.

This simple structure is analogous to Karl Popper's model of the continuous development of scientific knowledge (Popper 1972). Popper was concerned with the logic of scientific discovery, and in particular how it was not possible to prove by induction that an empirical theory is true because although all existing observations may be consistent with the theory, future observations that falsify the theory cannot be ruled out. Popper suggested that scientific knowledge does not proceed by the revelation of true theories (laws) as observations accumulate, but by a cycle of formulation, testing, falsification and reformulation of theories such that predictions are made with increasing accuracy and precision:

$$P_1 \rightarrow TS \rightarrow EE \rightarrow P_2$$

$P_1$  is the initial problem; TS is a trial solution to the problem; EE is error elimination, in which the trial solution is evaluated by observation; and  $P_2$  is a situation of increased knowledge. Knowledge development is continuous as  $P_2$  is an initial problem for which new trial solutions are proposed and evaluated. An important part of this concept is that observation cannot be prior to theory, as one must have a theory in order to decide what to observe (Popper

<sup>1</sup> In this paper, the term non-target organism refers to non-pest species. Pest species that are not intended to be controlled by the crop are "non-target pests". Protection of non-target pests is not an objective of the risk assessment, but they are important as a route of exposure of non-target organisms to transgenic proteins.

1972). Hence, science begins with a problem, not an attempt to solve the problem, and the sources of scientific problems are attempts to solve prior problems.

### Problem Formulation

An important consequence of Popper's conceptual model for scientific risk assessment is that the assessment should begin by defining the problem ( $P_1$ ), not by collecting data (EE). Problem definition normally begins with the objectives of the laws under which the proposed action is regulated; these "management objectives" are usually general statements about protection of the environment, although endangered species legislation may specify species and habitats to be protected.

The management objectives are not deducible scientifically; they are set by public policy, which will be based on political, economic, social and ethical, as well as scientific, criteria (Wolt & Peterson 2000; Johnson *et al.* 2007). To allow a scientific determination of risk, specific targets for protection, called assessment endpoints, must be derived from the management objectives. The assessment should comprise an entity (*e.g.*, a population of a particular species in a particular area) and a property of that entity (*e.g.*, the population size) (Newman 1998). For example in the UK, the management objective of conserving biodiversity is represented by an assessment endpoint of an index of the population sizes of bird species common on farmland (Gregory *et al.* 2004).

Usually it is not possible or desirable to measure directly the risk of a proposed action to the assessment endpoints. Instead a conceptual model that links the proposed action to the assessment endpoint is developed, and from this model specific "risk hypotheses" are derived. These hypotheses correspond to the trial solution part of Popper's model.

Because scientific knowledge derives from tests of hypotheses, not from proofs of hypotheses, it is not possible to prove that an action presents no risk to the assessment endpoints. It is possible, however, to attain high confidence that an action presents low risk ("is safe") by rigorous tests of risk hypotheses. For example, a conceptual model may suggest that the use of a chemical presents low risk to the abundance of an endangered species because the species will not be exposed the chemical. The risk hypothesis derived from this model is that the concentration of the chemical in the habitat of the endangered species is not significantly different from zero (or from a value that is "in effect zero" for the purposes of assessing risk). The hypothesis could be tested by mathematical modelling of the dispersal of the chemical under the proposed use, and confidence in the risk assessment could be increased by making conservative assumptions about the values of parameters in the model. If the risk hypothesis is not falsified after testing under highly

conservative conditions, there is high confidence that the use of the chemical presents negligible risk to the endangered species (Raybould 2006).

The derivation of risk hypotheses is called problem formulation, and is an essential, but often neglected aspect of risk assessments for the cultivation of GM crops. By erecting specific hypotheses to be tested, problem formulation identifies requirements for data. Without hypothesis testing, there is no method to identify data requirements because risk assessment will proceed on the flawed assumption that safety can be proved by the accumulation of data that show no effect. Collection of additional data could always be justified because it would provide “more evidence to prove safety”. Hypothesis testing provides a clear criterion to judge the value of additional data: unless the additional data offer a more rigorous test of the risk hypothesis than existing data, and thereby increase certainty of the risk assessment, they are superfluous (Raybould 2006).

If the introduction of environmentally beneficial products is delayed while superfluous data are collected, environmental risk is increased (Cross 1996). If delay can increase risk, no study can be free from risk, and requirements for data to assess the risk of an action must be balanced with the loss of potential benefits of that action while the data are collected. Therefore to minimise environmental risk, problem formulation should devise risk hypotheses that can be rigorously tested with the minimum need to acquire additional data (Raybould 2006).

### **Risk Characterisation**

The testing of risk hypotheses is called risk characterisation, and corresponds to the error elimination part of Popper’s scheme. Hypotheses are tested by comparing their predictions with observations. For a hypothesis to be scientific it must be possible to falsify it; a hypothesis that predicts every possible outcome of a test is not scientific (Popper 1959). It follows that good scientific theories make specific predictions, and rigorous tests of theories attempt to create conditions under which the theory is most likely to fail.

The logic of risk characterisation under Popper’s model is that a specific hypothesis should be formulated such that if it is not falsified, further risk characterisation would be unproductive. To build confidence that risk characterisation can stop, tests of the hypothesis should create conditions under which the hypothesis is most likely to fail. If the hypothesis is not falsified under those conditions, testing can stop and the risk assessment be completed. Hence risk assessment should seek to assess risk initially under “worst-case” conditions, and if the risk is minimal, no further data should be required. If the risk hypothesis is falsified, a new hypothesis is formulated ( $P_2$  under Popper’s scheme) and further characterisation of the risk is made under more realistic conditions. This is known as “tiered” risk assessment (Touart and

Maciorowski 1997).

Popper’s conceptual model shows that the development of scientific knowledge is continuous; knowledge acquired after a trial solution and error elimination presents new problems for which trial solutions are proposed. The same applies to risk assessment. No amount of corroborative data can prove a risk hypothesis. Also, new information may falsify theories on which the initial problem formulation was based, and therefore a different risk hypothesis should be tested to give sufficient certainty that the proposed action poses no unacceptable risks. The best that risk assessment can achieve is high confidence of minimal risk given present knowledge. The decision to stop risk characterisation is therefore a judgement that further testing will not increase knowledge of risk significantly, and hence effort is better spent increasing knowledge of a different problem.

### **Decision making**

Characterization of risk is not a decision to permit or forbid a proposed action. The results of the risk assessment must be evaluated along with any societal concerns that fall outside the risk assessment; this evaluation is risk analysis. Confusion between risk assessment and risk analysis is part of the reason for controversy about role of science in making decisions about the use of new technology. Non-scientific concerns about scientific advances have become confounded with scientific estimates of risk. This leads to “debates” about science, when what is really being debated is the weight that should be given to scientific assessments relative to other concerns about public policy when making decisions (Johnson *et al.* 2007). It is important to remember that risk assessment is led by policy, because the assessment endpoints are derived from management objectives set by policy, but risk characterization is not the only factor that determines decisions based on that policy.

### **Problem Formulation for Risk Assessments of GM Crops**

The preceding discussion argued that risk assessment should be viewed as conforming to the model of the continuous development of scientific knowledge proposed by Popper. As such, a risk assessment should begin with a problem and search for answers to that problem by testing hypotheses. Regarding a risk assessment as hypothesis testing recognises that safety cannot be proved, but can be indicated by rigorous tests of hypotheses that predict low risk. A risk assessment should not begin by collecting data and then try to work out whether they indicate a problem; this approach uses the flawed model of induction under which truths are supposed to emerge from accumulating observations. In the following sections, I will suggest some general hypotheses that can be tested to demonstrate low risk

from the cultivation GM crops with high confidence.

Most GM risk assessments are done to comply with laws, and the management objectives of these laws guide the risk assessment. For environmental risk assessments, the management objectives are often vague. In the United States, pesticidal proteins produced in GM plants are regarded as pesticides and therefore are regulated under the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), which seeks to

“...protect the public health and environment from the misuse of pesticides by regulating the labelling and registration of pesticides and by considering the costs and benefits of their use.”

In the European Union, GM crops for commercial cultivation are regulated under Directive 2001/18/EC, which requires that risk assessments

“...identify and evaluate potential adverse effects of the GMO, either direct [or] indirect, immediate or delayed, on human health and the environment which the deliberate release or placing on the market of GMOs may have.”

Both laws seek to protect the environment; however, “environmental protection” is too vague a concept to be analysed scientifically. A scientific risk assessment requires that the concept of environmental protection is made operational by deriving assessment endpoints. A common assessment endpoint for the risk assessment of GM crops is the abundance of non-target organisms; the risk to this endpoint is considered in this paper. The other common assessment endpoints are crop quality and yield, which are derived from management objectives of plant protection laws; the risks to these endpoints are considered elsewhere (Raybould 2005; 2007).

A simple and effective conceptual model to link the cultivation of a GM crop to harm to the abundance of non-target organisms is that non-target organisms could be reduced by exposure to toxic substances in the GM plant. The model makes two important assumptions. First, reductions in the abundance of predators and parasitoids solely due to control of the target pest are not considered harmful; control of pests is an intended effect of agriculture, and any method of control may have the effect of reducing the abundance of species that prey on or parasitise pests. Secondly, the effects of non-GM counterparts of GM crops on non-target organisms are acceptable. These assumptions greatly simplify the risk assessment as only differences in the composition the GM and non-GM crop need to be assessed for their effects on non-target organisms.

Under this model, the first task of the risk assessment is to characterise the differences between the GM crop and a non-GM counterpart. This plant

characterisation can be expressed as a risk hypothesis:

**Risk hypothesis 1:** there are no ecotoxicologically significant differences between the composition of the GM crop and the composition of its non-GM counterparts

If there are no significant differences between the composition of the GM crop and non-GM counterparts, the GM crop can be considered safe.

Some transgenic crops can be efficacious without expressing transgenic proteins; for example, virus resistance can be conferred by expression of transgenic RNA molecules without translation into protein (*e.g.*, Baulcombe 1996). Nevertheless, most GM plants are designed to express transgenic proteins, and so a minimum difference between the GM crop and its non-GM comparator is usually the presence of the transgenic protein, or possibly the concentration of the transgenic protein if the crop is designed to over-express a native plant protein.

Once the differences between the GM and non-GM crops are characterised, the next step in the risk assessment is usually to establish which organisms, if any, will be exposed to those differences. For simplicity, consider the presence of a transgenic protein in the GM crop to be the only difference identified. The concentration of the transgenic protein to which a non-target organism will be exposed as the result of cultivation of the GM crop is called the expected environmental concentration (EEC) (Table 1). If a non-target organism is not exposed to the protein this is equivalent to an EEC of zero, and if no non-target organisms are exposed to the transgenic protein, the crop can be considered safe. Again, this step can be expressed as a risk hypothesis:

**Risk hypothesis 2:** the expected environmental concentration EEC of the transgenic protein is not greater than zero for all non-target organisms<sup>2</sup>

For those organisms with an EEC greater than zero, the effect of that exposure should be evaluated.

Toxicological effects can be expressed as the concentration of a substance need to elicit a particular response; for example, the concentration of a substance that kills 50% of a group of test organisms, the median lethal concentration or LC<sub>50</sub>, is often used as means of comparing the toxicity of substances. For risk assessment, one may wish to know the highest concentration of a substance that elicits no adverse effect on an organism; this is the no observable adverse effect concentration or NOAEC. If no organism were

<sup>2</sup> The risk hypothesis is written as  $EEC \not> 0$ , not  $EEC = 0$ , because it is usual for statistical tests to test for no significant difference, not for equality.



**Table 1.** Generic estimates of expected environmental concentrations (EECs) for non-target organisms exposed to transgenic proteins *via* cultivation of GM crops.

Non-target organism type	Worst-case EEC	Realistic EEC	Comments
Above-ground arthropod	Highest mean concentration <sup>1</sup> of the transgenic protein in leaf tissue or pollen	0.2 X the worst-case EEC <sup>2</sup>	Several studies <sup>3</sup> have shown that the concentration of transgenic protein in most herbivores is <10% that of the GM crop on which they are feeding.
Soil invertebrate	Highest mean concentration of the transgenic protein in roots	Concentration of the transgenic protein if whole plants were ploughed into soil to a depth of 15cm immediately before harvest	The amount of transgenic protein per hectare is calculated by multiplying the whole plant concentration by the mass of a plant and the number of plants per hectare. The mass of one hectare of soil to a depth of 15 cm is <i>ca.</i> 2.25 x 10 <sup>9</sup> g.
Pollinator	Highest mean concentration of the transgenic protein in pollen	Average concentration of the transgenic protein in pollen multiplied by the proportion of the crop pollen in the diet	For maize, 50% pollen is a reasonably conservative assumption <sup>4</sup> ; for cotton, a higher proportion may be necessary.
Aquatic organisms	Highest mean concentration of the transgenic protein in pollen multiplied by the predicted worst-case density of pollen in a pond 2 m deep adjacent to field	As for worst-case, but use average concentration of protein and assume 50% loss of protein through degradation and transport and settling of pollen	Maize produces up to 315x10 <sup>9</sup> pollen grains per hectare during anthesis; 1 mg of maize pollen contains approximately 2,500 pollen grains <sup>5</sup> . Assume aquatic exposure to cotton is minimal as pollen is not dispersed by wind <sup>6</sup> .
Farmed fish	Concentration of protein detected in feed made from maximum proportion of grain that gives a nutritious diet	Multiply the worst-case EEC by the maximum predicted proportion of the product in the feed chain	Testing above the worst-case EEC is not possible if the worst-case EEC is estimated from the feed used in the hazard study.
Wild mammals	Worst-case daily dietary dose <i>via</i> a diet containing 100% seeds with highest mean concentration of transgenic protein	As for worst-case, but use average concentration of transgenic protein in seeds and realistic proportion of crop seed in diet	Formula for daily dietary dose is (FIR ÷ bw) x C: FIR is daily food intake rate, bw is body weight and C is the concentration of the transgenic protein in the diet. Estimates of FIR/bw for various mammalian diets are in the literature <sup>7</sup> .
Wild birds	As for wild mammals	As for wild mammals	Estimates of FIR/bw for various avian diets are in the literature <sup>7</sup> .

<sup>1</sup> The mean concentration of transgenic protein in GM plants at the developmental stage in which expression is highest

<sup>2</sup> Raybould *et al.* 2007

<sup>3</sup> *E.g.*, Head *et al.* 2001; Raps *et al.* 2001; Dutton *et al.* 2002; Howald *et al.* 2003; Obrist *et al.* 2005; Obrist *et al.* 2006a; Obrist *et al.* 2006b; Torres *et al.* 2006

<sup>4</sup> Babendreier *et al.* 2004

<sup>5</sup> Westgate *et al.* 2003; Depuis *et al.* 1987

<sup>6</sup> *E.g.*, Khan & Afzal 1950; Thies 1953; Sidhu & Singh 1961; Llewelyn & Fitt 1996

<sup>7</sup> Crocker *et al.* 2002

exposed to concentrations of a substance greater than its NOAEC, then that substance would pose minimal risk to non-target organisms. Thus a third hypothesis can be tested to determine the safety of exposure to a transgenic protein, or to any other potential toxin detected in a GM crop:

**Risk hypothesis 3:** the EEC of the transgenic protein is not greater than the NOAEC for all non-target organisms<sup>3</sup>

This is a very conservative risk hypothesis because it assumes that any adverse effect at concentrations below the EEC will give unacceptable effects in the field. In reality, density-dependent population dynamics and immigration mean even if there are adverse effects on populations the effects may be temporary; therefore, some other risk assessment methods, such as those used for chemical pesticides, test less conservative hypotheses such as that the EEC is not greater than 20% of the median lethal concentration (LC<sub>50</sub>) (US EPA 1998).

Rigorous tests of the above risk hypotheses provide a means to determine with high confidence (certainty) that a GM crop poses negligible risk to non-target organisms. The following sections discuss how rigour can be introduced into such tests.

### Hypothesis Testing for Risk Assessments of GM Crop

The following sections describe tests of the 3 risk hypotheses given above. To conform to the principles of tiered testing described earlier, tests of the hypotheses should be made under worst-case conditions if possible. If the hypothesis is corroborated under worst-case conditions, further testing should not be necessary as the risk can be characterised as minimal under all circumstances. As tests become more realistic, they become more specific to the conditions under which they are performed; hence tests under worst-case conditions should be relevant to all risk assessments, whereas tests under highly realistic conditions may be applicable to risk assessments for those conditions only. These are important considerations for risk assessors who are seeking data that are useful worldwide, not just in the region they were produced.

#### Plant characterization

The objective of plant characterization studies is to test the hypothesis that there are no ecotoxicologically significant differences between the GM crop and its non-transgenic counterparts. The

purpose of these studies is not to identify *any* difference between the GM and non-GM plants, but to identify differences in concentrations of substances that may have harmful effects on non-target organisms. Therefore plant characterization studies are targeted to particular substances; they should not attempt to compare global assessments of transcription or protein expression, or to assess metabolic profiles (*e.g.*, Baudo *et al.* 2006; Baker *et al.* 2006).

There are three main types of plant characterization study that are informative for non-target organism risk assessments: molecular characterization, compositional analysis and developmental studies. Molecular characterization uses methods such as Southern blotting and DNA sequencing to characterize the inserted DNA. Of particular importance is to test the hypothesis that inserted DNA will lead to the production of the intended transgenic proteins and will not lead to the production of unintended proteins. DNA sequencing can test for potential mutations and re-arrangements of the inserted DNA that may create new open-reading frames. If potential new open-reading frames are detected, further characterization such as Northern blotting may be needed to test whether unintended proteins are likely to be produced (*e.g.*, König *et al.* 2004).

Compositional analysis is mainly carried out to assess food safety (*e.g.*, Nair *et al.* 2002), but the data are relevant for non-target organism risk assessments. Compositional analysis tests the hypothesis that the GM crop and a non-GM near-isogenic line do not differ in compounds that are toxicologically relevant; such compounds include key nutrients, toxins, allergens, anti-nutrients, and other biologically active substances known to be associated with the crop (König *et al.* 2004). If differences are identified, the concentrations of the relevant substances should be compared with the natural range of variation in the crop; only if the concentration falls outside the natural range should assessment of the ecotoxicological impact of the difference be assessed.

The final element of plant characterization relevant to non-target organism risk assessment is the developmental expression study, which estimates concentrations of the transgenic proteins during growth of the GM crop. Protein concentrations are estimated by enzyme-linked immunosorbent assay (ELISA) (*e.g.*, Tijssen 1985). The concentration of transgenic proteins is measured in several tissues and, if relevant, at several developmental stages.

Usually, transgenic plants are intended to express new proteins, and therefore it may be expected that the risk hypothesis of no ecotoxicologically significant differences in the composition of the GM crop and its non-GM counterparts is inevitably false; and hence the developmental expression study is relevant for estimating EECs only (see below). This is not necessarily true; some transgenic plants are not

<sup>3</sup> Risk hypothesis 2 can be regarded as a special case of risk hypothesis 3; if  $EEC \geq 0$ ,  $EEC < NOAEC$ .

intended to produce protein, and have changed phenotypes mediated by RNA production without translation (*e.g.*, virus resistant crops discussed above). For these plants, a study analogous to the developmental study could be carried out to test of the hypothesis that no protein is translated from the transgene. If the protein was not detected, the hypothesis of no significant ecotoxicological difference would be corroborated and minimal risk to non-target organisms could be concluded without further testing.

Typically, the plant characterization phase shows that the only ecotoxicologically relevant difference between the GM and non-GM plants is the expression of the intended transgenic protein. If that is the case, the risk assessment should assess the potential effects of the transgenic proteins, while if other differences are detected, they should also be evaluated to assess the combined risk of the expression of the transgenic proteins and the differences in composition. The rest of the paper assumes that the transgenic proteins are the only ecotoxicologically relevant difference between the GM and non-GM plants; however, the principles described can also be used to evaluate other differences.

### Exposure

The exposure assessment estimates the EECs of the transgenic proteins and thereby identifies species potentially exposed to the proteins. The assessment can be regarded as a test of risk hypothesis 2, that non-target organisms are not exposed to the transgenic proteins. Should risk hypothesis 2 be falsified, the exposure assessment becomes part of a test of risk hypothesis 3, that the EEC is not greater than the NOAEC for all non-target organisms.

In addition to the results of the developmental expression study, several pieces of information are used to assess the environmental fate of the transgenic protein: the rate of its degradation in soil; the biology of the crop, particularly whether the crop forms self-sustaining populations outside agriculture; and the likelihood of gene flow from the crop to wild relatives.

The soil degradation study can be used to determine whether organisms that occur outside cultivation may be exposed to the transgenic protein *via* run-off in surface water. The study is also used to predict whether exposure to soil organisms may exceed exposure *via* plant tissue because of potential for accumulation of the transgenic protein in the soil.

The design of the soil degradation study is relatively simple. Soil is collected from the field and dosed with a test substance containing the transgenic protein; common test substances are lyophilized leaf tissue of the GM plant and microbially produced transgenic protein (see below). A negative control soil, collected and maintained in the same manner as the treatment soils, but without addition of test substance, is also used. The soils are incubated under conditions that sustain microbial activity. Soil samples are taken periodically and the activity of the transgenic protein is

estimated by mortality in a sensitive insect bioassay; negative control soil samples provide an estimate of background mortality and can be used to correct mortality of the treatment soil samples if necessary. Degradation of the transgenic protein is detected as a decrease in mortality of the sensitive insect in samples taken at increasing incubation times. The time for the activity of the protein to decline by 50%, the  $DT_{50}$ , is estimated from the rate of decline in mortality in the bioassay. Soil biomass and respiration may be measured at the beginning and the end of the study to test for healthy microbial activity (*e.g.*, Anderson & Domsch 1978); this is an important control as a long  $DT_{50}$  may be because of low microbial activity in the soil rather than inherent resistance of the protein to degradation.

Soil  $DT_{50}$ s have been estimated for many Cry proteins expressed in GM plants, with values between 2 and 22 days (US EPA 2001, US EPA 2003, US EPA 2005, US EPA 2007). These data predict that cultivation of GM plants containing these proteins is unlikely to lead to accumulation of transgenic proteins in the soil. This hypothesis has been corroborated by Head *et al.* (2002) and Dubelman *et al.* (2005), who showed that continuous cultivation of cotton containing Cry1Ac or maize containing Cry1Ab did not lead to the accumulation of the transgenic proteins in soil. These results were not surprising as most proteins do not persist or accumulate in soil because they are inherently degradable in soils that have healthy microbial activity (*e.g.*, Burns 1982, Marx *et al.* 2005, and references therein). In conclusion, if the soil  $DT_{50}$  of a transgenic protein is shown to be short, it can be concluded that organisms outside fields are unlikely to be exposed to the protein *via* run-off, and organisms inside fields are unlikely to be exposed to concentrations of the protein greater than those in the GM plant.

Other routes by which organisms outside fields in which GM crops are cultivated may be exposed to transgenic proteins are gene flow and the establishment of feral populations<sup>4</sup> of the GM crop. Often sufficient information about the biology of a crop is already available to show that exposure to transgenic proteins is unlikely; for example, in the United States transgenic maize is unlikely to hybridize with wild plants or to establish feral populations (US EPA 2001). In cases where the biology of the crop is less well-known, or if it is likely that the genetic modification could increase the likelihood of establishment of feral populations, new data may be required to test whether exposure *via* these routes is unlikely; these data requirements are discussed in a companion paper (Raybould 2007).

If environmental fate data indicate that gene flow, feral populations and soil accumulation are unlikely,

---

<sup>4</sup> Self-sustaining populations of crops outside cultivation.

the hypothesis that there is no exposure to the transgenic protein is corroborated for organisms that are not exposed to the crop. Exposure of non-target organisms to the transgenic protein *via* the crop may occur directly through consumption of crop tissue, or by consumption of prey that has eaten crop tissue. The main groups of organisms exposed *via* this route are terrestrial arthropods that are predators of crop pests, soil invertebrates and aquatic organisms that may be exposed to pollen deposited in surface water. Animals that eat the crop are generally regarded as pests, not non-target organisms; however, wild birds and wild mammals that consume the crop are often regarded as non-target organisms in the environmental risk assessment. Farm animals potentially exposed to the transgenic protein *via* feed are not usually included in the environmental risk assessment, although farmed fish are sometimes included as they are not generally included in risk assessments for food and feed.

The data from the developmental expression study are used to calculate EECs for the above groups of organisms. A useful method is to calculate a “worst-case” exposure, where the diet of the non-target organism is 100% the relevant tissue of the GM crop, and a “realistic” exposure, where the transgenic protein

is diluted in the prey, in the soil or by other means relevant to the organism. The realistic EEC still provides a conservative estimate of exposure as it assumes all individuals of a species are exposed. Worst-case exposures may be used when the objective of the risk assessment is protection of individual animals, such as members of endangered species, and realistic exposures may be used when the objective is the protection of populations or ecological function (Raybould *et al.* 2007). The methods for calculating worst-case and realistic exposures are given in Table 2 and most are derived from the US EPA (2001) and Raybould *et al.* (2007).

### Hazard

If the exposure assessment indicates that non-target organisms may be exposed to the transgenic protein (*i.e.*, the EEC is greater than zero) hazard data are required to test the third risk hypothesis: the NOAEC of the transgenic protein is not less than the EEC. This hypothesis can be expressed as test of whether the ratio of the EEC to the NOAEC is less than 1 for all NTOs; EEC/NOAEC is termed the hazard quotient (HQ) (Kelly & Roy-Harrison 1998).

**Table 2.** A typical set of non-target organisms for testing the hazard of a transgenic protein (based on species tested for the risk assessment of MIR604 maize expressing modified Cry3A for control of corn rootworm (Raybould *et al.* 2007)).

Functional Group	Test species	Common name	Order: Family
Above-ground arthropod	<i>Coccinella septempunctata</i>	Seven-spot ladybird	Coleoptera: Coccinellidae
Above-ground arthropod	<i>Orius insidiosus</i>	Insidious flower bug	Hemiptera: Anthocoridae
Soil invertebrate	<i>Poecilus cupreus</i>	Ground beetle	Coleoptera: Carabidae
Soil invertebrate	<i>Aleochara bilineata</i>	Rove beetle	Coleoptera: Staphylinidae
Soil invertebrate	<i>Eisenia foetida</i>	Earthworm	Haplotaxida: Lumbricidae
Pollinator	<i>Apis mellifera</i>	Honeybee	Hymenoptera: Apidae
Aquatic organism	<i>Oncorhynchus mykiss</i>	Rainbow trout	Salmoniformes: Salmonidae
Aquatic organism	<i>Daphnia magna</i> <sup>1</sup>	Water flea	Cladocera: Daphniidae
Wild mammal	<i>Mus musculus</i>	Mouse	Rodentia: Muridae
Wild bird	<i>Colinus virginianus</i>	Bobwhite quail	Galliformes: Phasianidae

<sup>1</sup> Species not tested for mCry3A risk assessment because of low expression in MIR604 pollen. Included for illustration.



For some proteins it may be possible to conclude that the NOAEC is greater than the EEC by knowledge of the mode of action of the protein, or from data on prior exposure. For example, herbicide tolerance in GM crops is often conferred by proteins that have high homology with native plant proteins or that are members of classes of proteins that are ubiquitous (*e.g.*, acetyltransferases); therefore, there is high confidence that there will be no adverse effects of these proteins to wildlife<sup>5</sup> at concentrations found in GM crops. Consequently, specific studies to assess the ecological hazard of proteins conferring herbicide tolerance are usually not required (*e.g.*, Peterson & Sharma 2005; Garcia-Alonso *et al.* 2006). Although the spectrum of activity of proteins used to confer insect resistance in GM crops is often well-known (*e.g.*, Schnepf *et al.* 1998), there is less confidence in the conclusion of no adverse effects of these proteins on non-target organisms when expressed in GM plants, and therefore specific hazard studies have been required for these proteins.

To provide a rigorous test of the hypothesis  $EEC/NOAEC \leq 1$ , hazard studies should increase the likelihood of detecting an adverse effect of the transgenic protein at a given concentration. Laboratory studies provide a higher likelihood than field studies of detecting an effect because extraneous variation can be minimised so increasing the power to detect an effect (*e.g.*, Maund *et al.* 1997, Rand & Zeeman 1998, de Jong *et al.* 2005). Additional rigour is provided by laboratory studies because they offer the possibility of exposing species to concentrations of the transgenic proteins in excess of the EEC; uncontained field studies are limited to exposures at the EEC. Exposures in excess of the EEC are useful for extrapolation to species that may be more sensitive to the transgenic protein, and for extrapolation to longer exposures to the transgenic protein that may be encountered in the field compared with the laboratory. Test species selection and study designs are designed to minimize the need to extrapolate to more sensitive species or long exposures.

It is not possible to obtain estimates of NOAECs for all non-target organisms that may be exposed to the transgenic proteins; organisms representative of functional or taxonomic groups likely to be exposed are tested and the data are used to make predictions about the sensitivity of similar species. If certain species are likely to be more sensitive to the transgenic protein, and a robust test method is available, they should be chosen as representatives of their group as they provide the best estimate of the minimum

NOAEC: this gives the most rigorous available test of the risk hypothesis and minimizes the need to extrapolate for species sensitivity. For example, in hazard studies of modified Cry3A expressed in MIR604 maize, Raybould *et al.* (2007) selected 3 species of beetle for non-target arthropod testing because the intended target pests are chrysomelid beetles (corn rootworm; *Diabrotica virgifera virgifera* and *D. barberi*). A typical set of test species for functional groups often exposed to transgenic proteins *via* GM crops is given in Table 2.

The choice of species for hazard testing must be pragmatic; species should only be used for regulatory studies if a robust test method is available. Essential requirements for a robust test method are low mortality and normal development in the negative control groups, and exposure to the test substance in the treatment groups. Protocols for testing the effects of pesticides (*e.g.*, US EPA 1996, Candolfi *et al.* 2000) can provide useful guidelines for testing transgenic proteins; these include sample sizes, statistical power, maximum control mortality, minimum positive control mortality, and the environmental conditions under which the test should be maintained. Many pesticide test protocols use acute exposure to the test substance *via* contact, whereas hazard testing of transgenic proteins may require long-term dietary exposure; therefore substantial method development may be required to adapt pesticide test protocols for testing proteins (*e.g.* Duan *et al.* 2006; Raybould *et al.* 2007). Often the most difficult aspect of method development is identification of an artificial diet that will allow development of the test species for a substantial part of its life-cycle, while preserving bioactivity of the transgenic protein. Cooking the diet in a microwave oven to denature proteases before addition of the protein test substance reduces the likelihood of loss of bioactivity.

Exposure to transgenic proteins in laboratory hazard studies is usually *via* microbial test substances incorporated into diet. Bacteria, such as *Escherichia coli*, are transformed with the gene used to create the GM plants and used to produce large quantities of the transgenic protein by fermentation. The advantage of microbial test substances over plant test substances is that exposure in the hazard studies can exceed the EEC by two or three orders of magnitude if necessary. In theory, purified protein could be obtained from transgenic plants, but enormous numbers of plants would be required to obtain the quantities of protein produced by microbial fermentation. To ensure that the transgenic protein in the microbial test substance is a suitable surrogate for the protein in the plant, several tests are carried out, including comparisons of molecular weight, glycosylation, cross-reactivity with antibodies and bioactivity against a sensitive insect pest (*e.g.*, Raybould *et al.* 2007); DNA or protein sequences of the genes in the bacterial expression system and in the transgenic plant may also be

---

<sup>5</sup> The term “wildlife” is used instead of “non-target organism”, as there is no “target” organism of proteins conferring herbicide tolerance.

compared.

Exposure to the protein in hazard studies is usually designed to be a low multiple of the worst-case EEC. The multiple of the EEC used in study is called the margin of exposure<sup>6</sup>. A margin of exposure of about 10 (10X EEC) is regarded by many as sufficient to extrapolate results from tested species to the species for which they are surrogates and so provide protection for all potentially exposed non-target organisms (*e.g.*, US EPA 2007). Higher concentrations of protein can be used if very low HQs are required to provide confidence in the risk assessment (see below). In studies that use artificial diets, aliquots of treated diet can be kept frozen and freshly thawed samples supplied daily to the test organisms to help ensure that exposure to bioactive protein is maintained throughout the study.

The responses measured in a hazard study (the test endpoints) should reveal effects that are potentially relevant ecologically, not seek to detect *any* difference that may exist between the groups exposed to the transgenic protein and the negative control groups. In studies of invertebrates, larval development, adult emergence and reproduction are considered to be sensitive, but ecologically relevant endpoints; in studies of vertebrates, weight gain, feeding behaviour and mortality are common endpoints. If there are no statistically significant differences between the treatment and negative control groups in the test endpoints, it can be concluded that the NOAEC<sup>7</sup> is at least the concentration of transgenic protein present in the test.

For studies that use artificial diets, it is important to confirm that the protein was present at the nominal concentration; it is not usually necessary to confirm exposures in studies that supply a single dose of protein by oral gavage, or protein in aqueous solutions

<sup>6</sup> The margin of exposure (MoE) is not the same as a safety factor. For example, if one is testing the risk hypothesis that  $EEC/NOAEC \leq 1$ , hazard testing at the EEC (MoE = 1) may be sufficient to indicate low risk. An MoE of 10 provides additional corroboration of the risk hypothesis and increases the confidence in the risk assessment; however, it is not essential that testing is done at 10X EEC, and any effects observed at 10X EEC would not indicate an unacceptable risk, provided they were not observed at 1X EEC. Application of a 10-fold safety factor in effect means that the risk hypothesis that indicates acceptable risk is  $EEC/NOAEC \leq 0.1$ . To test this hypothesis, there must be an MoE of at least 10X EEC, and adverse effects at this concentration would indicate unacceptable risk requiring further evaluation (see US EPA 2007, for an excellent discussion of the relationship between MoEs and safety factors).

<sup>7</sup> Or no observable adverse effect level (NOAEL) when exposure is *via* a single dose of protein.

that are replaced regularly. The transgenic protein is extracted from aliquots of treated diet kept frozen for the duration of the exposure phase of the study. The concentration of the protein is measured by ELISA and a Western blot is used to confirm that the ELISA is measuring intact protein, not degradation products (*e.g.*, Raybould *et al.* 2007). Bioactivity of the protein can be confirmed using sensitive insect bioassays of thawed aliquots of treated diet (Duan *et al.* 2006, Raybould *et al.* 2007). If the ELISA, Western blot and bioassay indicate little degradation of the transgenic protein, it can be concluded that the protein was present in the freshly thawed diet at the nominal concentration for the duration of the test, and that the  $NOAEC \geq$  nominal concentration<sup>8</sup>. A positive control treatment, in which a known orally active toxin is incorporated into the diet, is sometimes used to corroborate exposure in the protein-treated group.

Hazard testing concludes the initial data collection phase of the risk assessment. The data are used to estimate risk by testing the risk hypotheses given above. This phase of the risk assessment is risk characterization.

### Risk characterization

If the risk hypotheses are corroborated by tests of sufficient rigour, it may be concluded with high confidence that the GM crop poses low risk to non-target organisms. For GM crops expressing proteins for insect resistance, with no other detectable ecotoxicologically relevant differences from a suitable conventional crop comparator, the key hypothesis under test is that the  $HQ \leq 1$ . This hypothesis is tested by comparing exposure data (EEC estimates) with hazard data (estimates of the NOAEC).

A series of HQs is obtained for species that represent groups of organisms potentially exposed to the protein. If the HQs are all below 1, then low risk is indicated to the species tested; confidence that the risk is low to *all* potentially exposed NTOs can be derived from the rigour with which the hypothesis  $HQ \leq 1$  was tested. If all HQs are well below 1 (say  $< 0.1$ ) using worst-case estimates of the EEC, the risk hypothesis is corroborated under highly rigorous conditions, giving confidence of low risk to all NTOs; on the other hand, if the HQs are all  $\approx 1$  using the realistic EEC, the hypothesis is less rigorously corroborated and confidence is lower that risk is low to all NTOs. However, it should be remembered that even the realistic EECs are conservative because they assume all individuals are exposed, and HQs are maxima if the NOAEC has been derived from a study using a single concentration of protein. Therefore an HQ of 1 based on a realistic EEC may be considered a rigorous test of

<sup>8</sup> The NOAEC may be higher than the nominal concentration in the study, but further studies at higher concentrations would be needed to establish that.

the hypothesis of no adverse effects of the transgenic in the field (e.g., US EPA 2007).

The characterization of risk does not constitute a decision, it simply makes explicit to decision makers the risk hypothesis under test and the rigour with which the hypothesis has been tested. The decision whether corroboration of the risk hypothesis has been made with sufficient rigour to permit cultivation of the GM crop is part of risk analysis and may include information other than the risk assessment (e.g., Wolt & Peterson 2000, Johnson *et al.* 2007). Two regulators may come to different decisions from the same set of HQ values depending on their interpretation of the policies and regulations under which they are working: one may decide that enough information has been collected to make a decision with sufficient confidence, while the other may decide that further testing is required.

If further testing is required, the tests should increase the rigour with which the risk hypothesis is tested. For the hypothesis  $HQ \leq 1$ , increased rigour could involve hazard testing at higher concentrations of protein or testing additional species; in both cases, if the hypothesis was corroborated the confidence of low risk to all non-target organisms is increased. In general, if no effect has been seen in a hazard study at concentrations of at least 1X EEC, a field study will not increase the rigour with which the hypothesis  $HQ \leq 1$  is tested because uncontrolled variation makes detection of an effect more difficult than in the laboratory. The realism of a test is not necessarily an indication of the usefulness of a study for decision-making; the crucial attribute of a test is the rigour with which it tests the risk hypothesis. Hence requests for further data should be predicated on increasing the rigour of tests, and hence increasing confidence in the risk characterization, not on increasing the amount of data *per se* (Raybould 2006).

### Higher Tier Tests

Testing may falsify the risk hypothesis  $HQ \leq 1$  for certain groups of organism. In these cases, new risk hypotheses can be created and tested. For example, the hypothesis that the toxic effect of the transgenic protein will not significantly decrease the population size of the organism can be tested using data on the toxicity of the protein under more realistic conditions, a more precise estimate of exposure of the organism to the protein, or both. The realism of the studies can be increased up to large-scale field studies. Studies that increase the realism of the testing are called “higher tier” studies in contrast to the unrealistically conservative “tier 1” studies described above.

An example of higher tier testing and risk characterization is the work that characterized the risk of maize expressing Cry1Ab to monarch butterflies (*Danaus plexippus*). Cry1Ab is active against

Lepidoptera and is expressed in maize primarily to control European corn borer (ECB; *Ostrinia nubilalis*). The monarch is potentially exposed to Cry1Ab via maize pollen settling on the leave of its food plant (common milkweed, *Asclepias syriaca*). Laboratory studies (Hellmich *et al.* 2001) indicated adverse effects on the development of monarch larvae from exposure to maize pollen at densities found on some milkweed plants in the field (Pleasants *et al.* 2001); hence the hypothesis that  $HQ \leq 1$  was falsified for monarchs exposed to Cry1Ab maize, at least under worst-case EECs.

Further work demonstrated that less than 1% of the US and Canadian monarch population was likely to be exposed to toxic concentrations of Cry1Ab (Sears *et al.* 2001). Exposure characterization showed that most milkweed populations occurred sufficiently far from maize fields that pollen deposition would be negligible, and that the monarch larvae feeding did not coincide with maize anthesis. The predicted low exposure to Cry1Ab, the possible reduced exposure of monarchs to insecticides used to control ECB, and the rapid recovery of monarch populations from catastrophic events such as frost in their winter roosting habitat (e.g., Calvert *et al.* 1983) indicated that the risk to monarchs from cultivation of Cry1Ab maize was low (US EPA 2001).

The conclusions from higher tier studies tend to be more specific than lower tier studies; for example, the data on monarch exposure to maize pollen are not generally applicable to all Lepidoptera because of differences in distribution. The results of tier 1 studies, on the other hand, are generally applicable; laboratory toxicity studies and worst-case exposure estimates indicate with high certainty that Cry1Ab maize is unlikely to be toxic to any species of beetle, regardless of where it occurs. Therefore unless tier 1 studies are impractical, higher tier studies should be considered only when a powerful risk hypothesis has been falsified by lower tier data.

### Stacked traits

Many new GM crops will be “breeding stacks” – combinations of traits brought together by conventional breeding. If the traits have gained regulatory approvals separately, what are suitable risk hypotheses for assessing the risks of stacks?

A simple risk hypothesis is that the effect of the mixture of transgenic proteins is not greater than the addition of the effects of the proteins separately; in other words if the concentrations of the proteins were approximately equal in the stack, and all HQs for the proteins separately were  $\leq 0.5$ , the HQ for mixture will be  $\leq 1$ . One way to test this hypothesis<sup>9</sup> is to treat the

<sup>9</sup> Usually, testing of this hypothesis is only required for stacks that combine two or more insect resistance traits.

mixture of proteins as a new active ingredient and carry out exposure and hazard estimates as described for the single proteins; however, this is inefficient as simpler methods are available that test the risk hypothesis with greater rigour.

A simple method for testing that exposure to the proteins is not greater than additive is to compare expression of the proteins in the stack with expression in the relevant single trait GM plant. If expression in several tissues at several developmental stages is not significantly higher in the stack, then the hypothesis of no greater than additive exposure is corroborated with confidence.

The most powerful test of the hypothesis that the hazard is not greater than additive (not synergistic) is to examine the effects of the mixture in species that are sensitive to at least one of the proteins; it is unlikely species insensitive to the proteins are more likely to detect synergism than are sensitive species. Pest species are usually used as the sensitive bioassay species as they are often highly sensitive to at least one of the proteins (*e.g.*, they are the target pest or closely related to it taxonomically) and can be conveniently reared in the laboratory.

For combinations of two proteins, test designs differ depending on the sensitivity of available pest species. If a species sensitive to both proteins is available, dose response curves for the separate proteins would be obtained. The predicted response of the species to mixtures of the proteins can be obtained from these data.

The predicted effect depends upon the modes of action of the proteins. If the proteins have similar modes of action, the predicted  $LC_{50}$  of the mixture can be estimated from the  $LC_{50}$  of the proteins separately, using a model called simple similar action. If the proportions of protein A and protein B in the mixture are  $r_A$  and  $r_B$ , respectively, and their respective  $LC_{50}$ s when tested separately are  $LC_{50(A)}$  and  $LC_{50(B)}$ , the predicted  $LC_{50}$  is given by the harmonic mean of the separate  $LC_{50}$ s, weighted by the proportion of each protein in the mixture (Tabashnik 1992):

$$LC_{50(mixture)} = \frac{1}{\left[ \frac{r_{(A)}}{LC_{50(A)}} + \frac{r_{(B)}}{LC_{50(B)}} \right]}$$

If the predicted  $LC_{50(mixture)}$  is statistically significantly lower than the observed  $LC_{50(mixture)}$ , the hypothesis of no synergism is falsified.

If the proteins have different modes of action, the predicted effect of the mixture should be calculated using a model called independent joint action. Under this model, if a certain amount of protein A alone kills

$x\%$  of a sample, and a certain amount of protein B kills  $y\%$ , the predicted percentage kill of a mixture of these amounts of protein is given by  $x + y - (xy/100)$  Colby (1967)<sup>10</sup>. The observed and expected mortalities are compared over a range of concentrations. There is no test of statistical significance; the predicted dose response curves are compared with the expected dose response curves and if there is greater mortality than expected over the range of concentrations the hypothesis of synergism is falsified.

For pairs of proteins that target different pest species, a simple experimental design uses analysis of variance to test the effect of the presence of the “non-toxic” protein on the toxicity of the other protein. Consider protein A, toxic to species X and non-toxic to species Y, and protein B, toxic to species Y and non-toxic to species X. The first part of a test for absence of synergism is to obtain dose response curves to estimate the  $LC_{30}$ ,  $LC_{70}$  and  $LC_{90}$  for the proteins against their respective target species. Then two separate experiments are set up with the same design:

Bioassay 1 with species X, comprising 4 treatments (controls not shown)

1.  $LC_{30}$  of protein A
2.  $LC_{70}$  of protein A
3.  $LC_{30}$  of protein A +  $LC_{90}$  of protein B to species Y
4.  $LC_{30}$  of protein A +  $LC_{90}$  of protein B to species Y

Bioassay 2 with species Y, comprising 4 treatments (controls not shown)

1.  $LC_{30}$  of protein B
2.  $LC_{70}$  of protein B
3.  $LC_{30}$  of protein B +  $LC_{90}$  of protein A to species X
4.  $LC_{30}$  of protein B +  $LC_{90}$  of protein B to species X

The mortality of the bioassay species is assessed in each treatment. The data are subject to 2-way ANOVA to test the hypotheses of no effect of the concentration of the toxin, and no effect of the non-toxin on the response to the toxin. A statistically significant effect of the non-toxin indicates non-additivity of the toxicity of the mixture, and if mortality is greater in the presence of the non-toxin, the hypothesis of no synergism is falsified.

If the studies corroborate the hypothesis of no synergism in sensitive pest species, it is likely that there will be no synergism of the mixture against non-target organisms, and that the risk hypothesis of  $HQ \leq 1$  for all non-target organisms exposed to the mixture is corroborated. No testing of the mixture against non-

<sup>10</sup> If protein A kills  $x\%$ , protein B will kill  $y\%$  of the remainder, *i.e.*  $x + y/100(100-x) = x + y - xy/100$ .



target organisms should be necessary under these circumstances. Predicting the effects of mixtures of more than 3 proteins can be complex (*e.g.*, Cassee *et al.* 1998), and although in theory tests for lack of synergism in pest species are more sensitive, hazard tests of the mixture of proteins to non-target organisms may be a more tractable approach for risk assessments of 3 or more proteins. A full complement of tests such as illustrated in Table 2 should not be necessary to establish lack of synergism; tests on species most closely related to a target pest of one of the proteins, or on species with particularly high predicted exposure should provide a sufficient test of the risk hypothesis.

### Conclusions

Environmental risk assessments for GM plants should be viewed as tests of risk hypotheses not collections of data. Good problem formulation should identify phenomena that are necessary for the GM plant to adversely affect the targets for protection (the assessment endpoints), and it follows that rigorous risk hypotheses, that is those that are most informative for decision-making, are those that predict the absence of those phenomena.

Confidence in the risk assessment is provided by the rigour with which the risk hypotheses are tested. Where possible, testing should begin under conditions most likely to reveal that the risk hypothesis is false. If the risk hypothesis is corroborated under those conditions, there can be confidence that the risks posed by the GM plant are low.

Risk hypotheses are often tested most rigorously under laboratory conditions because the potential effects of the GM plant can be amplified and isolated from most other sources of variation. If risk hypotheses are corroborated under laboratory conditions, the temptation to supplement the risk assessment with field studies should be avoided. First, field studies will not add to confidence in the conclusion of no risk because their power to falsify the risk hypotheses is lower than the laboratory studies. Secondly, collection of additional data introduces a source of environmental risk because it may delay or prevent the introduction of an environmentally beneficial product.

Delay comes from the collection of the data, and also from the extra time required by decision-makers to evaluate the data. Extra data may confuse rather than clarify risk leading to the unwarranted rejection of an application for a registration of a GM crop, or conversely to an approval of a product that has a high probability of causing environmental harm. Collection of data also increases the development costs of GM crops, and if costs become prohibitive, potentially beneficial products may not be developed; this is a particular problem for public sector institutions in developing countries (*e.g.*, Cohen 2005), although large multinational companies are also affected as research and development budgets are not unlimited.

In summary, data should only be collected for risk assessment purposes if it provides a more rigorous test of a powerful risk hypothesis than is available with existing data. Collection of unnecessary data should thereby be minimised, and environmental risks decreased as the introduction of environmentally beneficial products is not delayed unduly.

### References

- Anderson, J.P.E. and K.H. Domsch. 1978. A physiological method for quantitative measurement of microbial biomass in soils. *Soil Biol. Biochem.* 10: 215-221.
- Babendreier, D., N. Kalberer, J. Romeis, P. Fluri and F. Bigler. 2004. Pollen consumption in honey bee larvae: a step forward in the risk assessment of transgenic plants. *Apidol.* 35, 293-300.
- Baker, J.M., N.D. Hawkins, J.L. Ward, A. Lovegrove, J.A. Napier, P.R. Shewry and M.H. Beale. 2006. A metabolic study of substantial equivalence of field-grown genetically modified wheat. *Plant Biotech. J.* 4: 381-392.
- Baudo, M.M., R. Lyons, S. Powers, G.M. Pastori, K.J. Edwards, M.J. Holdsworth and P.R. Shewry. 2006. Transgenesis has less impact on the transcriptome of conventional wheat than conventional breeding. *Plant Biotech. J.* 4: 369-380.
- Baulcombe, D. 1996. Mechanisms of pathogen-derived resistance to viruses in transgenic plants. *The Plant Cell* 8: 1833-1844.
- Burns, R.G. 1982. Enzyme activity in soil: location and possible role in microbial ecology. *Soil Biol. and Biochem.* 14: 423-427.
- Calvert, W.H., W. Zuchowski and L.P. Brower. 1983. The effect of rain, snow and freezing temperatures on overwintering monarch butterflies in Mexico. *Biotropica* 15: 42-47.
- Candolfi, M.P., S. Blümel, R. Forster, F.M. Bakker, C. Grimm, S.A. Hassan, U. Heimbach, M.A. Mead-Briggs, B. Reber, R. Schmuck and H. Vogt. 2000. Guidelines to Evaluate Side-effects of Plant Protection to Non-target Arthropods. International Organization for Biological and Integrated Control of Noxious Animals and Plants, West Palearctic Regional Section (IOBC/WPRS). Gent.
- Cassee, F.R., J.P. Groten, J.P. van Bladeren and V.J. Feron. 1998. Toxicological evaluation and risk assessment of chemical mixtures. *Crit. Rev. Toxicol.* 28: 73-101.
- Cohen, J.I. 2005. Poorer nations turn to publicly developed GM crops. *Nature Biotech.* 23: 27-33.
- Colby, S.R. 1967. Calculating synergistic and antagonistic responses of herbicide combinations. *Weeds* 15: 20-22.
- Crocker, D., A. Hart, J. Gurney and C. McCoy. 2002. Project PN0908: Methods for Estimating Daily Food Intake of Wild Birds and Mammals. Central Science Laboratory. Report to the UK Department



- of Environment, Food and Rural Affairs. Available online at: <[http://www.pesticides.gov.uk/uploadedfiles/Web\\_Assets/PSD/Research\\_PN0908.pdf](http://www.pesticides.gov.uk/uploadedfiles/Web_Assets/PSD/Research_PN0908.pdf)>. Accessed 30th May, 2007.
- Cross, F.B. 1996. Paradoxical perils of the precautionary principle. *Wash. L. Law Rev.* 53: 851-925.
- De Jong, F.M.W., B.J.W.G. Mensink, C. Els Smit and M.H.M.M. Montforts. 2005. Evaluation of ecotoxicological field studies for authorization of plant protection products in Europe. *Hum. Ecol. Risk Assess.* 11: 1157-1176.
- Depuis, I., P. Roeckel, E. Matthys-Rochon and C. Dumas. 1987. Procedure to isolate viable sperm cells from corn (*Zea mays* L.) pollen grains. *Plant Physiol.* 85: 876-878.
- Duan, J.J., M.S. Paradise, J.G. Lundgren, J.T. Bookout, C. Jiang and W.N. Wiedenmann. 2006. Assessing nontarget impacts of Bt corn resistant to corn rootworms: tier-1 testing with larvae of *Poecilus chalcites* (Coleoptera: Carabidae). *Environ. Entomol.* 35: 135-142.
- Dubelman, S., B.R. Ayden, B.M. Bader, C.R. Brown and C. Jiang and D. Vlachos. 2005. Cry1Ab protein does not persist in soil after 3 years of sustained Bt corn use. *Environ. Entomol.* 34: 915-921.
- Dutton, A., H. Klein, J. Romeis, F. Bigler. 2002. Uptake of Bt-toxin by herbivores feeding on transgenic maize and consequences for the predator *Chrysoperla carnea*. *Ecol. Entomol.* 27: 441-447.
- Garcia-Alonso, M., E. Jacobs, A. Raybould, T.E. Nickson, P. Sowig, H. Willekens, P. van der Kouwe, R. Layton, F. Amijee, A.M. Fuentes and F. Tencalla. 2006. A tiered system for assessing the risk of genetically modified plants to non-target organisms. *Environ. Biosafety Res.* 5: 57-65.
- Gregory, R.D., D.G. Noble and J. Custance. 2004. The state of play of farmland birds: population trends and conservation status of lowland farmland birds in the United Kingdom. *Ibis* 146 (suppl. 2): 1-13.
- Head, G., C.R. Brown, M.E. Groth, and J.J. Duan. 2001. Cry1Ab protein levels in phytophagous insects feeding on transgenic corn: implications for secondary exposure risk assessment. *Entomol. Exp. Appl.* 99: 37-45.
- Head, G., J.B. Surber, J.A. Watson, J.W. Martin and J.J. Duan. 2002. No detection of Cry1Ac protein in soil after multiple years of Bt cotton (Bollgard) use. *Environmental Entomology* 31: 30-36.
- Hellmich, R.L., B.D. Siegfried, M.K. Sears, D.E. Stanley-Horn, M.J. Daniels, H.R. Mattila, T. Spencer, K.G. Bidne and L.C. Lewis. 2001. Monarch larvae sensitivity to *Bacillus thuringiensis* purified proteins and pollen. *Proc. Natl. Acad. Sci. USA* 98: 11925-11930.
- Hill, R.A. and C. Sendashonga. 2003. General principles for risk Assessment of living modified organisms: lessons from chemical risk assessment. *Environ. Biosafety Res.* 2: 81-88.
- Howald, R., C. Zwahlen and W. Nentwig. 2003. Evaluation of Bt oilseed rape on the non-target herbivore *Athalia rosae*. *Entomol. Exp. Appl.* 106, 87-93.
- Johnson, K.L., A.F. Raybould, M.D. Hudson and G.M. Poppy. 2007. How does scientific risk assessment of GM crops fit within the wider risk analysis? *Trends Plant Sci.* 12: 1-5.
- Kelly, E.J. and W. Roy-Harrison. 1998. A mathematical construct for ecological risk: a useful framework for assessments. *Human Ecol. Risk Assessment* 4: 229-241.
- Khan, A.H. and M. Afzal. 1950. Natural crossing in cotton in Western Punjab IV. Agents of natural crossing. *Agron. J.* 42: 236-238.
- König, A., A. Cockburn, R.W.R. Crevel, E. Debruyne, R. Grafstroem, U. Hammerling, I. Kimber, I. Knudsen, H.A. Kuiper, A.A.C.M. Peijnenburg, A.H. Penninks, M. Poulsen, M. Schauzu and J.M. Wal. 2004. Assessment of the safety of food derived from genetically modified (GM) crops. *Food Chem. Toxicol.* 42: 1047-1088.
- Llewelyn, D. and G. Fitt. 1996. Pollen dispersal from two field trials of transgenic cotton in the Namoi Valley, Australia. *Mol. Breeding* 2: 157-166.
- Marx, M-C., E. Kandeler, M. Wood, N. Wermbter and S.C. Jarvis. 2005. Exploring the enzymatic landscape: distribution and kinetics of hydrolytic enzymes in soil particle-size fractions. *Soil Biology and Biochemistry* 37: 35-48.
- Maund, S.J., T.N. Sherratt, T. Stickland, J. Biggs, P. Williams, N. Shillabeer and P.C. Jepson. 1997. Ecological considerations in pesticide risk assessment for aquatic ecosystems. *Pesticide Sci.* 49: 185-190.
- Nair, R.S., R.L. Fuchs and S.A. Schuette. 2002. Current methods for assessing safety of genetically modified crops and exemplified by data on Roundup Ready soybeans. *Toxicol. Path.* 30: 117-125.
- Newman, N.C. 1998. *Fundamentals of Ecotoxicology*. Ann Arbor Press, Chelsea, MI.
- Obrist, L.B., A. Dutton, R. Albajes and F. Bigler. 2006a. Exposure of arthropod predators to Cry1Ab toxin in Bt maize fields. *Ecol. Entomol.* 31: 143-154.
- Obrist, L.B., A. Dutton, J. Romeis and F. Bigler. 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, L.B., H. Klein, A. Dutton, A. and F. Bigler. 2005. Effects of Bt maize on *Frankliniella tenuicornis* and exposure of thrips predators to prey-mediated Bt toxin. *Entomol. Exp. Appl.* 115: 409-416.
- Patton, D.E. 1998. Environmental risk assessment: tasks and obligations. *Hum. Ecol. Risk Assess.* 4: 657-670.

- Peterson, R.K.D and L.M. Sharma. 2005. A comparative risk assessment of genetically engineered, mutagenic, and conventional wheat production systems. *Transgenic Res.* 14: 859-875.
- Pleasants, J.M., R.L. Hellmich, G.P. Dively, M.K. Sears, D.E. Stanley-Horn, H.R. Mattila, J.E. Foster, P. Clark and G.D. Jones. 2001. Corn pollen deposition on milkweeds in and near cornfields. *Proc. Natl. Acad. Sci.* 98: 11919-11924.
- Popper, K.R. 1959. *The Logic of Scientific Discovery.* Hutchinson, London.
- Popper, K.R. 1972. *Objective Knowledge: an Evolutionary Approach.* Oxford University Press.
- Power, M. and S.M. Adams. 1997. Perspectives of the scientific community on the status of ecological risk assessment. *Environ. Manag.* 21: 803-830.
- Raps, A., J. Kehr, P. Gugerli, W.J. Moar, F. Bigler and A. Hilbeck. 2001. Immunological analysis of phloem sap of *Bacillus thuringiensis* corn and of the nontarget herbivore *Rhopalosiphum padi* (Homoptera: Aphididae) for the presence of Cry1Ab. *Mol. Ecol.* 10: 525-533.
- Rand, G.M. and M.G. Zeeman. 1998. Ecological risk assessment approaches within the regulatory framework. *Hum. Ecol. Risk Assess.* 4: 853-886.
- Raybould, A. 2005. Assessing the environmental risks of transgenic volunteer weeds. In: J. Gressel (ed.) *Crop Fertility and Volunteerism.* Boca Raton: CRC Press.
- 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. II. Assessing the consequences of gene flow. *Bioassay* (submitted).
- Raybould, A., D. Stacey, D. Vlachos, G. Graser, X. Li and R. Joseph. 2007. Non-target organism risk assessment of MIR604 maize expressing mCry3A for control of corn rootworm. *J. Appl. Ent.* 131: 391-399.
- Schnepf, E., N. Crickmore, J. van Rie, D. Lereclus, J. Baum, J. Feitelson, D.R. Zeigler and D.H. Dean. 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol. Mol. Biol. Rev.* 62: 775-806.
- Sears, M.K., R.L. Hellmich, D.E. Stanley-Horn, K.S. Oberhauser, J.M. Pleasants, H.R. Mattila, B.D. Siegfried and G.P. Dively. 2001. Impact of Bt corn pollen on monarch butterfly populations: a risk assessment. *Proc. Natl. Acad. Sci.* 98: 11937-11942.
- Sidhu, A.S. and S. Singh. 1961. Studies on agents of cross pollination in cotton. *Indian Cotton Growing Rev.* 15: 341-353.
- Stern, P.C. and H.V. Fineberg. 1996. *Understanding Risk: Informing Decisions in a Democratic Society.* National Academy Press, Washington, DC.
- Tabashnik, B. 1992. Evaluation of synergism among *Bacillus thuringiensis* toxins. *Appl. Environ. Microbiol.* 58: 3343-3346.
- Thies, S.A. 1953. Agents concerned with natural crossing of cotton in Oklahoma. *Agron. J.* 45: 481-484.
- Tijssen, P. 1985. *Practice and Theory of Enzyme Immunoassays.* Elsevier Science Publishers, Amsterdam.
- Torres, J.B., J.R. Ruberson and M.J. Adang. 2006. Expression of *Bacillus thuringiensis* Cry1Ac protein in cotton plants, acquisition by pests and predators: a tritrophic analysis. *Agric. Forest Entomol.* 8: 191-202.
- Touart, L.W. and A.F. Maciorowski. 1997. Information needs for pesticide registration in the United States. *Ecol. Appl.* 7: 1086-1093.
- (US EPA) United States Environmental Protection Agency. 1996. Microbial Pesticide Test Guidelines OPPTS 885.4340 Nontarget Insect Testing, Tier I. Available online at: <[http://www.epa.gov/opptsfrs/publications/OPPTS\\_Harmonized/885\\_Microbial\\_Pesticide\\_Test\\_Guidelines/Series/885-4340.pdf](http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/885_Microbial_Pesticide_Test_Guidelines/Series/885-4340.pdf)>. Accessed 30th May, 2007.
- (US EPA) United States Environmental Protection Agency. 1998. Guidelines for ecological risk assessment. *Fed. Regist.* 63: 26846-26924.
- (US EPA) United States Environmental Protection Agency. 2001. Biopesticides Registration Action Document – *Bacillus thuringiensis* Plant Incorporated Protectants. Available online at: <[http://www.epa.gov/pesticides/biopesticides/pips/bt\\_brad2/3-ecological.pdf](http://www.epa.gov/pesticides/biopesticides/pips/bt_brad2/3-ecological.pdf)>. Accessed 30th May, 2007.
- (US EPA) United States Environmental Protection Agency. 2003. Biopesticides Registration Action Document: Event MON863 *Bacillus thuringiensis* Cry3Bb1 Corn. Available online at: <[http://www.epa.gov/oppbppd1/biopesticides/ingredients/tech\\_docs/cry3bb1/2\\_c\\_cry3bb1\\_environl.pdf](http://www.epa.gov/oppbppd1/biopesticides/ingredients/tech_docs/cry3bb1/2_c_cry3bb1_environl.pdf)>. Accessed 30th May, 2007.
- (US EPA) United States Environmental Protection Agency. 2005. Biopesticides Registration Action Document: *Bacillus thuringiensis* Cry34Ab1 and Cry35Ab1 Proteins and the Genetic Material Necessary for their Production (Plasmid Insert PHP 17662) in Event DAS-59122-7 Corn. Available online at: <[http://www.epa.gov/oppbppd1/biopesticides/ingredients/tech\\_docs/brad\\_006490.pdf](http://www.epa.gov/oppbppd1/biopesticides/ingredients/tech_docs/brad_006490.pdf)>. Accessed 30th May, 2007.
- (US EPA) United States Environmental Protection Agency. 2007. Biopesticides Registration Action Document: Modified Cry3A Protein and the Genetic Material Necessary for its Production (Via Elements of pZM26) in Event MIR604 Corn SYN-IR604-8. Available online at: <[http://www.epa.gov/oppbppd1/biopesticides/ingredients/tech\\_docs/brad\\_006509.pdf](http://www.epa.gov/oppbppd1/biopesticides/ingredients/tech_docs/brad_006509.pdf)>. Accessed 30th May, 2007.
- Westgate, M.E., J. Lizaso and W. Batchelor. 2003. Quantitative relationships between pollen shed density and grain yield in maize. *Crop Sci.* 43: 934-942.
- Wolt, J.D. and R.K.D. Peterson. 2000. Agricultural biotechnology and societal decision-making: the role of risk analysis. *AgBioForum* 3: 39-46.