Environ. Biosafety Res. 7 (2008) 1–9
© ISBR, EDP Sciences, 2008
DOI: 10.1051/ebr:2008001

Available online at:
www.ebr-journal.org

Review article

Identification of potentially hazardous human gene products in GMO risk assessment

Hans BERGMANS[1]*, Colin LOGIE[2], Kees VAN MAANEN[3], Harm HERMSEN[4], Michelle MEREDYTH[5] and Cécile VAN DER VLUGT[1]

[1] GMO Office, National Institute for Public Health and the Environment, Bilthoven, The Netherlands
[2] Nijmegen Department of Molecular Biology, Nijmegen Center for Molecular Life Sciences, Radboud University, Nijmegen, The Netherlands
[3] Animal Health Service, Deventer, The Netherlands
[4] Centre for Biological Medicines and Medical Technology, National Institute for Public Health and the Environment, Bilthoven, The Netherlands
[5] Departments of Genomics, Biotech and Internal Medicine I, Technical University Dresden, Germany

Genetically modified organisms (GMOs), *e.g.* viral vectors, could threaten the environment if by their release they spread hazardous gene products. Even in contained use, to prevent adverse consequences, viral vectors carrying genes from mammals or humans should be especially scrutinized as to whether gene products that they synthesize could be hazardous in their new context. Examples of such potentially hazardous gene products (PHGPs) are: protein toxins, products of dominant alleles that have a role in hereditary diseases, gene products and sequences involved in genome rearrangements, gene products involved in immunomodulation or with an endocrine function, gene products involved in apoptosis, activated proto-oncogenes. For contained use of a GMO that carries a construct encoding a PHGP, the precautionary principle dictates that safety measures should be applied on a "worst case" basis, until the risks of the specific case have been assessed. The potential hazard of cloned genes can be estimated before empirical data on the actual GMO become available. Preliminary data may be used to focus hazard identification and risk assessment. Both predictive and empirical data may also help to identify what further information is needed to assess the risk of the GMO. A two-step approach, whereby a PHGP is evaluated for its conceptual dangers, then checked by data bank searches, is delineated here.

Keywords: genetic modification / genetic engineering / genetically modified organism / GMO / contained use / risk assessment / hazard identification / GMO regulation / viral vector / human gene products

INTRODUCTION

The biosafety of genetically modified organisms (GMOs) was first discussed around 1975, acknowledging that by genetic modification organisms could acquire new traits that had never before been observed in that genetic context. The new traits could cause new phenotypes that had never existed before. It was envisaged that some new phenotypes could be hazardous, as they could cause harm to the organism itself or to the environment (Berg et al., 1975). To estimate how dangerous it would be to deal with a GMO, methods for risk assessment were devised (*e.g.* EU, 2002). Most methods start by identifying what potential hazards expression of the newly acquired genes in the GMO might pose. However, a working definition of a 'hazardous' gene product has never been precisely described.

The NIH Guidelines (latest version: NIH, 2002), on which GMO regulations for contained use have been based in many countries, do not define a hazardous gene product, but make clear that hazard is interpreted in terms of the potential for causing disease. The OECD 'Blue Book' (OECD, 1986), which is the basis for all science based methodologies for GMO risk assessment developed to date, speaks of 'conjectural hazards', *i.e.* hazards 'not based on incident', which could be phrased in colloquial language as hazards based on an 'educated guess'. According to the Blue Book, 'when recombinant DNA techniques were first introduced there was a natural concern as to their potential hazards, but after more than a

* Corresponding author: hans.bergmans@rivm.nl

decade of experimentation under controlled conditions, these hazards have remained conjectural and not based on incident. [...] Any potential hazards of [...] the use of recombinant DNA organisms are expected to be of the same nature as for other biological agents, namely: infection hazard - the potential for disease in man, animals and plants following exposure to the living organism [...]; the toxic, allergenic or other biological effect of the organism or cell, its components or its naturally occurring metabolic products; the toxic, allergenic or other biological effect of the product synthesized by the organism; [and] effects for agricultural and environmental applications.'

Directive 98/81/EC (EU, 1998) of the European Union (EU), on the environmental risk assessment of contained use of GMOs, offers only a general explanation of what would be considered as hazards of GMOs: 'The following should be considered as potentially harmful effects: disease to humans including allergenic and toxic effects; disease to animals or plants; deleterious effects due to the impossibility of treating a disease or providing an effective prophylaxis; deleterious effects due to establishment or dissemination into the environment; deleterious effects due to natural transfer of inserted genetic material to other organisms.' In the guidance notes to Directive 98/81/EC (EU, 2000), the concept 'potentially harmful effect' is defined in similarly general terms, *i.e.* 'those effects which may give rise to disease, render prophylaxis or treatment ineffective, promote establishment and/or dissemination in the environment which gives rise to harmful effects on organisms or natural populations present or harmful effects arising from gene transfer to other organisms.' EU Directive 2001/18/EC (EU, 2001), on the environmental risk assessment of deliberate release of GMOs, provides a similar list in the description of the first step of environmental risk assessment in its Annex II.

The concept of hazardous gene products has been further developed on a case by case basis in the hazard identification step of the risk assessment of deliberate release of GMOs, mainly for genetically modified crop plants. Summaries of these risk assessments are available, for instance on the web site of the Biosafety Clearing House of the Cartagena Protocol on Biosafety[1].

For the risk assessment of contained use of GMOs, however, there is less specific guidance available on what 'hazardous gene products' are. Two classes of gene products, toxins and virulence factors, are generally considered as hazardous. In practice, however, regulators are frequently faced with risk assessments for the use of GMOs where the products of the cloned gene are not so straightforwardly classified as 'hazardous' or 'non-hazardous'. This study intends to show an approach to identifying 'potentially hazardous gene products' (PHGPs): gene products that may be hazardous when synthesized in a new physiological background, and/or at levels or times that differ from their original physiological background. The approach is illustrated for the risk assessment of contained use of GM viral vectors encoding human or mammalian genes, as an example. This example is chosen because of its relevance for current regulatory practice, as well as because this is a clear model case. The potential risk caused by viral vectors carrying genes that encode PHGPs is that they may infect a worker, *e.g.* by aerosols or by needle stick accidents, and cause adverse health effects. The study aims to spark off further discussions on how to consider PHGPs in general GMO risk assessment, and in contained use particularly.

HAZARD IDENTIFICATION

The first step, identifying potentially harmful effects of the GMO, is crucial for the outcome of the risk assessment process. When the risk is assessed during the developmental stages of a GMO, only limited empirical knowledge is available. This is typically the case in the risk assessment of contained use.

In practice, hazard identification in these cases is highly 'conjectural', *i.e.* based on 'educated guesses'. There is however some firm ground on which the potentially harmful effects of a GMO can be predicted from the combined knowledge about the host organism and the donor trait(s) introduced into that recipient by genetic modification. If a vector is used in the modification process, its properties are also to be taken into account. Based on this knowledge and experience, a scenario can be developed to predict the effect of the inserted genetic information on the properties of the recipient organism, and the consequences of these effects for the interaction of the GMO with its expected receiving environment. In most cases this scenario is expected to be rather straightforward, as it will be expected that the donor trait will be expressed in the same way in the GMO as it is in the donor organism. However, this does not have to be the case. The gene product(s) of the inserted gene(s) may have different interactions in the physiological background of the recipient. While a gene product may have intrinsic hazardous, *e.g.* toxic, properties, its actual hazard depends very much on the genetic and physiological context of the host organism that was used for the genetic modification. Moreover, the inserted genes may be expressed differently, depending on the activity of the promoter and other regulatory sequences that govern their expression. Consequently, conclusions on the hazardous nature of a gene product can only be made for a specific

[1] http://bch.biodiv.org/decisions/riskassessments.shtml

genetic and physiological context, under specific conditions of use.

For this study, the relevant parameters setting the context and conditions are: the contained use of human genes and their products, introduced by the most commonly used viral vector systems, adenoviral and lentiviral vectors, in eukaryotic cells, tissues or organisms.

The viral vectors

In most GMO applications that involve viral vectors, 'safe' vector systems are used. For adenoviral vectors, this means that the vector is rendered replication defective by deleting one or more essential genes, and the vector production system has been designed so that only viral vector particles, and no wild-type replication-competent Adenovirus (RCA) is formed, *e.g.* the Per.C6 host cell line-vector system as described by Fallaux et al. (1998). Preventing the emergence of RCA reduces the chance that the intrinsically replication-deficient viral vector will replicate through complementation. For the same reason, the lentiviral vectors production systems have been constructed in such a way that the chance of emergence of a replicationcompetent lentivirus (RCL) is negligible, *e.g.* the 'third generation' system described by Dull et al. (1998).

From the point of view of biosafety, there is one crucial difference between adenoviral vectors and lentiviral vectors. An accidental infection with a replication-deficient adenovirus will lead only to a temporary expression of transgenes on a vector borne by the virus. The infected cells carrying the vector will eventually be cleared by the immune system. After an accidental infection with a lentivirus-borne vector, however, the vector may integrate into the genome of the infected cell, which may lead to permanent expression of the transgenes located on the vector.

Function of the insert within the genetic context of the host cell

Within the context of this study, the potential hazard of the gene product will depend on the function of the inserted gene in the background of the cell/organism. The inserted gene may be identical to the other copies of the gene that are already present in the organism, or it may be a mutant allele that could influence the cell's physiology if its function is dominant. The situation is more complicated if more than one gene is inserted, which could lead to unexpected effects, especially when a high multiplicity of infection is used. As this is not yet common practice, experience is lacking. Still, already with viral vectors carrying only one inserted gene, unexpected results have been observed that are probably based upon effects of a transgene product and one or more resident gene products, *e.g.* the case of an Ectromelia vector carrying an IL-4 gene that caused very severe effects in mice (Jackson et al., 2001). In reaction to this observation, the UK Health and Safety Executive (HSE, 2002) published guidelines on risk assessment of genetically modified viral vectors that can alter immune responses, arguing that special attention should be given to cases where the genetic modification would modulate or circumvent host defenses. On the part on the inserted genes, the guidelines point to genes that encode immune modulators such as cytokines, molecules that bind cytokines or interfere with antigen presentation, fusion to molecules that enhance antigen presentation, costimulatory molecules, chemokines and chemokine receptors as well as growth factors.

This study does not deal with situations where the added gene(s) complement a function that has been lost, *e.g.* in the case of a hereditary disease. Such added genes will in principle not have a harmful effect, provided that spatial and temporal expression levels mimic wild type levels. This study also does not deal with the creation of a knock-out mutant by inserting a disrupting sequence into a gene; the hazards involved will not be different from the hazards of creating the same mutant by mutagenesis.

Expression of the insert and processing of the gene product

The regulatory context of the gene, which determines its level of expression, will differ from case to case, according to the circumstances of the experiment. In many cases the inserted gene is expressed from a promoter that is active in all or most cell and tissue types, like the Cytomegalovirus (CMV) promoter, which has been used extensively. Its level of expression is constitutively high, so the gene product is abundant. This level may be higher than the expression of the gene under normal physiological circumstances, and the expression pattern may deviate appreciably from its normal pattern, which may be restricted to only one cell or tissue type. If this expanded expression pattern is expected to have adverse effects, those effects could be mitigated by using conditional promoters, *e.g.* tissue-specific or inducible promoters.

After synthesis, the gene product may not be immediately active. Activity may depend on post-translational modification, *e.g.* glycosylation, acetylation, phosphorylation, as well as transport of the protein to a specific cell compartment or secretion from the cell, and processing, *e.g.* by proteolytic cleavage.

For the sake of simplicity it is assumed here that transgenes are expressed from a promoter such as the CMV

promoter, *i.e.* expressed constitutively in all cell types. They are assumed to be active in the cell that they are produced in, and may also be secreted from the cell into the extracellular medium, or into the bloodstream if the GMO has infected a human or animal.

Conditions of use

In practice, adenoviral and lentiviral vectors are handled under contained use, in laboratories or in animal facilities. Typical activities are: inoculating of cell cultures with a virus, handling cell cultures in which the vectors are replicating and infectious vector particles are being formed, harvesting and processing cell-free media containing infectious vector particles, inoculating laboratory animals with cells or cell-free media containing infectious vector particles. The main hazard of these activities is infection of the worker with the viral vector, potentially followed by the worker shedding the viral vector into the environment and infecting other people.

For adenoviral vectors, infection could occur orally, especially from hand to mouth, by exposure to aerosols, or by needle stick accidents. The exposure of a worker to a replication-defective adenoviral vector may lead to a transient infection, and infected cells will be degraded by the immune system.

For HIV-based lentiviral vectors, the most likely route of infection is by needle stick accidents. If an infection with a lentiviral vector occurs, there would be a high chance that some of the infected cells could be permanently transformed by integration of the lentiviral vector into the genome, since lentiviral vectors are designed for just this purpose. Similar to adenovirus-infected cells, lentivirus-infected cells will be degraded by the immune system if the gene product is not isogenic to the worker.

For both adenoviral and lentiviral vectors, the level of exposure will depend mainly on the laboratory practice, including the number of vector particles involved, which may be high when high titer vector suspensions are being used. This will be the case only during production and direct handling of batches of the vector, as opposed to handling infected cultures or animals for experimental purposes under conditions when the vector cannot replicate.

Viral vectors carrying genes coding for PHGPs can be tested on laboratory animals for potential dangers of the PHGPs. In such experiments, the PHGPs pose a biohazard only if the vector is shed by the animals. Their potential harmful effect on the animals is not seen as a biosafety issue. Such experiments may however raise ethical concerns, and the appropriate use of laboratory animals in such experiments will be considered by specific Institutional Animal Care and Use Committees (IACUCs). Because of these considerations, IACUCs are also in need of guidelines for animal experiments involving transgenic animals or the use of viral vectors such as discussed in this study.

CONCEPTS OF THE IDENTIFICATION OF PHGPS

In general, the hazards involved in cloning any human (or mammalian) gene have been thought to be minimal. Indeed, it is hard to envisage that the expression of a 'normal' human gene under 'normal' circumstances could lead to a hazardous situation, and this is probably a suitable point of departure for the consideration of 'hazardous' gene products originating from human genes. If a human gene is expressed in such a way that homeostasis in the human body cannot be affected, nothing much will go wrong. When homeostasis is affected, however, a hazardous situation could develop. *A priori* it is expected that such conditions could occur if expression occurs in cellular compartments or in tissues where it normally does not, if the gene product is made in its normal environment but at levels far beyond those that occur naturally, or if gene expression is mistimed either within the cell cycle or during development. Under these circumstances, seemingly innocuous human gene products could be suspected of having toxic activity.

From inheritance studies, some alleles are known to have gene products involved in pathogenesis; in particular dominant alleles are of interest here. More gene products could conceivably be involved, *e.g.* mutations that are lethal in normal development could accidentally be constructed. Gene products could, for instance, produce unexpected adverse effects, such as genome rearrangements, which may be oncogenic. Furthermore, overproduction of gene products has been postulated (*e.g.* ACGM, 2000) as a conjectural hazard, leading to autoimmune disease or storage disorders. On the other hand, the potential effects of overproduction could be counteracted by increased catabolism of the gene product.

When these considerations are taken into account, a number of gene products may be seen as potentially hazardous in a specific context: gene products that have toxic properties; allergens - defined as immunologically active biomolecules that have a potential to affect persons with a constitutive histamine hypersensitivity, immunomodulators, gene products with an endocrine function, products of dominant alleles that have an etiological role in hereditary diseases and disorders, gene products and sequences involved in genome rearrangements, gene products involved in apoptosis or oncogenesis. Finally, we note that currently 474 miRNAs have been confirmed (Mazière and Enright, 2007). The miRNAs that impinge on the protein-coding genes or their regulators listed above could indeed be considered as PHGPs.

The use of databases for determining the potentially hazardous nature of human gene products

In the previous section, PHGPs were identified on the basis of arguments that are highly conjectural. This type of argumentation is useful at the start of a discussion on potential hazardous effects of a gene product, but for a more unambiguous identification of a gene product as either being or not being a PHGP, empirical data are necessary. A thorough screening of the available scientific literature, using adequate search engines and search strategies, will show whether relevant data are available; a negative result of a screening may be taken as evidence that no hazardous effects have been observed, although this type of negative conclusions should be drawn with caution.

In practice, a main problem is sorting through and interpreting the relevance of the large number of references that are found. One practical way to solve this is to restrict the search to reviews, with the rationale that important trends will be covered in this way, while the number of 'hits' will be substantially lower.

The problem of obtaining an authoritative overview may be solved at least to a certain extent if an interpretative literature database is available that contains data compiled and annotated with scientific rigor by a number of editors who are experts in the field. The Online Mendelian Inheritance in Man (OMIM) database is a high quality database by these standards[2]. OMIM contains critical comments on published data, complete with references, as well as extensive links to MEDLINE and sequence records in the Entrez system, and links to additional related resources. The OMIM database is accessible through a search engine with extensive advanced search possibilities, and would appear to be well suited for data mining of information on whether a human gene product may be seen as a PHGPs.

In order to provide some examples of the use of literature databases, the OMIM database was screened (results refer to a screening performed in December 2006) for information underpinning the potential hazardous nature of three groups of human gene products that may be seen as PHGPs: gene products that may have toxic properties, gene products active in genome rearrangement, gene products related to pathogenesis, and hazards of overexpression of gene products.

There are only a few examples of human proteins or peptides that are seen as toxins. 'Lymphotoxins' (OMIM record *153440) are immunomodulators that have a cytotoxic function. 'Anaphylatoxins', *e.g.* the C3a (OMIM record *605246), C4a and C5a factors of the complement system. These are polypeptide fragments of larger proteins with cytotoxic properties, which cause severe effects when overproduced. These proteins have functions in large physiological processes, but they clearly can provoke toxic effects under certain circumstances, and may be seen as PHGPs. Human gene products may also form structures that are comparable to structures formed by bacterial toxins. The mutant amyloid and synuclein proteins (OMIM *163890) in patients with familial Alzheimer's or Parkinson's disease, for instance, resemble bacterial pore-forming toxins.

The results indicate that differential expression of a gene appears to be an important factor for turning a gene product into a PHGP. Mutation and evolution is a second factor. In OMIM *606110, describing LY6/neurotoxin 1 (LYNX1), a homolog of snake alpha-neurotoxin, the structural and functional homology between components of the snake venom and nontoxic mammalian gene products is discussed. It appears that snake toxin genes may have evolved from recruited copies of the genes coding for 'normal' gene products (Fry, 2005).

The occurrence of genome rearrangements is known to have potential adverse effects, at least in specific cases, *e.g.* where genome rearrangements lead to dysregulation of specific genes and oncogenesis. Out-of-context expression of this type of genes would therefore be expected to turn their gene products into PHGPs. It is however not easy to find data in the OMIM database that support this supposition. Data on adverse effects of recombinases in the OMIM database are scarce. The OMIM database provides no data on the effects of out-of-context expression of general recombinases, *e.g.* the Rad51 gene, probably because these experiments have not been performed in a way that their effect on humans (or animals) could be tested. This notion is supported by the finding that overexpression of Rad51 in fruit flies has been found to be lethal and linked to induction of apoptosis (Yoo and McKee, 2004).

The human site-specific recombinases RAG1 and RAG2 (OMIM *179612) function in the production of the gene families encoding the different classes of antibodies and T cell receptor genes. RAG1 or RAG2 deficiency both lead to immune deficiency. The RAG proteins are only expressed in cells of the immune system. When expressed out of context, the gene products should be regarded as potentially hazardous, because they could cause genome rearrangements (Barreto et al., 2001).

Evaluation of what effect an active transposase gene could have is not straightforward. The OMIM database does not provide indications of adverse effects. However from the observations of Han et al. (2004), it could

[2] OMIM is a catalogue of human genes and genetic disorders authored and edited (and copyrighted) by Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere, and developed for the World Wide Web by NCBI, the National Center for Biotechnology Information. The URL for OMIM is http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db= OMIM

be conjectured that misexpression of active transposases present in human and animal genomes could be deleterious, if such an expression could be achieved. They show that transcription of the ORFs of the L1 transposon occurs at a very low level, and that this is an inherent effect of the mRNA, which apparently slows down transcription elongation. Overexpression of the L1 transposase appears not to occur, even if high mRNA levels are produced from an overexpression vector (Han et al., 2004).

The OMIM database specifically keeps track of alleles of human genes involved in pathogenesis, in a specific part of the database called 'morbid map', with currently around 5000 entries. For the context of this study, however, the entries on dominant alleles would be most relevant, because these could encode PHGPs, as a 'third copy' allele next to two most probably 'healthy' alleles. Also, a search for '(pathogen* or disorder*)' in the OMIM database yielded over 5000 hits for genes that have allelic forms that are involved in pathogenesis. The gene products encoded by these alleles should be considered potentially hazardous if they function in a dominant fashion, but this does not render their wild type 'healthy' counterparts directly suspect.

Overexpression on its own appears to be an important factor for deleterious effects. A search for 'overexpression' yielded nearly 1400 hits. Overexpression may for instance lead to unbalanced cell growth and tumor formation: examples are overexpression of a p53 binding protein homolog, MDM2 (OMIM *164785), of Myc (OMIM *190080), or of growth factors, *e.g.* VEGF, whose overexpression leads to highly malignant gliomas (OMIM *192240). Overexpression may also lead to cell death, *e.g.* PTEN, the phosphatase and tensin homolog (OMIM *601728). Overexpression of interleukins causes different effects: overexpression of IL-3 in mice leads to 'a motor neuron disease with several features of human ALS and progressive muscular atrophy' (OMIM *147740), but overexpression of IL-12 'may be useful in preventing UV-induced skin cancer' (OMIM *161561). Screening of the hits in searches for 'apoptosis' (1033 hits), 'apoptosis and regulat*' (744 hits), 'apoptosis and overexpression' (361 hits) and 'oncogen* and regulat*' (471 hits) or 'oncogen* and overexpression' (215 hits) supports the notion that deregulation and overexpression are major effectors. One example of deleterious results of transgenic (over)expression of TGFB1 (transforming growth factor beta-1) was found in OMIM *190180: overexpression of transgenic TGFB1 in rats caused severe fibrotic disease in the liver and fibrosis and glomerular kidney disease, depending on the level of expression. Searches in OMIM did not however support the notion, postulated earlier as a conjectural hazard, that products of genes involved in autoimmune disease or storage disorders can become PHGPs by overexpression. Searches for 'autoimmune and overexpression' storage disorders yielded 60 and 28 hits respectively, none of them showing a direct linkage of overexpression of a gene to a storage disorder.

Thus, the conclusion that overexpression of a gene product may lead to a hazardous situation appears to be warranted. This then leads to the question as to what in fact overexpression is. The level of expression of a transgene should be evaluated against its natural expression level, which will be different for each gene, and also for different tissues. The Human Anatomic Gene Expression Library (H-ANGEL) may help here. H-ANGEL (Tanino et al., 2005) is 'a resource for information concerning the anatomical distribution and expression of human gene transcripts. H-ANGEL utilizes categorized mRNA expression data from both publicly available and proprietary sources. H-ANGEL is accessible at http://www.jbirc.aist.go.jp/hinv/h-angel/.'

If 'misexpression', *e.g.* overexpression or expression in tissues where, or at a time in development when, the PHGP is normally not expressed, could have adverse effects, the use of conditional expression strategies, *i.e.* the use of inducible promoters, or promoters that confer distinct tissue and temporal specificity, could be considered as a hazard mitigating factor.

CONCLUSIONS

In risk assessment of GMOs, the first step is to identify hazards from transgene products expressed by the GMO. This paper tackles the problem, in this first step, of how to identify transgene products that could be hazardous even in the context of contained use. Only part of the problem is covered here: the context of human genes cloned in viral vectors that are considered as safe. To address the problem in a more comprehensive way other situations should also be considered: *e.g.* replicating viral vectors, such as Vaccinia, but also the more recently developed replication competent Adenovirus vectors, prokaryotic and eukaryotic microorganisms, plants and animals, and donor sequences derived from non-mammalian sources. It is still very likely that broad issues can be tackled using the approach we have outlined: (1) identifying the main scenarios whereby workers and/or the environment may be exposed; (2) identifying which classes of gene products could cause potential adverse effects through those main scenarios; (3) corroborating the actual occurrence of the potential adverse effects through searches of the available literature, with a focus on overexpression as a trigger.

It can be foreseen that besides helping with risk assessment, this approach will also identify major gaps in the available data, the baseline information that is a prerequisite for scientific risk assessment. This positive contribution to risk assessment in general could spark off new

areas of useful risk assessment research, *i.e.* research focused on gathering information that is useful for risk assessment ('need-to-know'), rather than data that are interesting ('nice-toknow'), but that are not helpful for the main questions of risk assessment.

In the practice of risk management, this means that identification of a gene product as a PHGP based on a hazard scenario alone may lead to enhanced biosafety measures only on the basis of the precautionary approach. Risk assessors should generally be able to consider gene products of human provenance as harmless in principle, until they have been defined as (potentially) harmful through good scientific argumentation, following the three steps mentioned above. Finally, the use of extrinsically inducible promoters, or promoters that have a (combination of) distinct tissue and temporal specificity, should be considered as a major hazard mitigating factor.

The identification of hazardous gene products in the risk assessment of the contained use of GMO applications can be divided into two stages. In the first stage, it is checked whether the gene product(s) of the GMO fits one of the conceptual scenarios, or whether another scenario could be devised that could lead to an adverse effect. Based on current considerations, the following gene products are seen as PHGPs: products of dominant alleles that have a role in hereditary diseases and disorders, gene products and sequences involved in genome rearrangement, products that are seen as toxins or that have toxic or allergenic properties, products involved in immunomodulation and in general products with an endocrine function, products involved in apoptosis or oncogenesis. The list is not intended to be comprehensive; the risk assessment of new GMOs may lead to other conceptual scenarios.

This type of scenarios that predict adverse effects typically invokes the precautionary principle, and consequent allocation of the GMO to a higher biosafety class (see Annex IV of EU (1998) for a description of containment measures in different biosafety classes). We recommend, however, that if the first stage of hazard identification leads to the conclusion that a gene product expressed by a GMO is a PHGP, this conclusion is checked further in a second stage of hazard identification. The examples presented in this study show that the use of high quality databases, complemented by bioinformatic approaches and expert opinions, can be useful to test whether the conjectural hazards can be corroborated by facts and demonstrated hazards. Although different gene products should be tested on a case-by-case basis, the results shown do not provide very much support for the current hazard scenarios. In fact, the only firm general conclusion that could be drawn is that overexpression, or in general misexpression, may lead to adverse effects. As science progresses, the information in the databases will keep growing, and the usefulness of bioinformatics for risk assessment may also be expected to increase. Risk assessors should actively follow these developments and apply the newly gained knowledge.

In order to draw a conclusion from data retrieved from databases, it is important to consider extensively the significance of the retrieved data by discussing the choice of database and search strategy. This is particularly important if the absence of data is used as an argument (*e.g.* 'there are no indications for toxicity of the gene product'). The absence of data can only be an argument if it can be made plausible that data, if they exist, would have been found by the search strategy, which should therefore be described in detail. The significance of the retrieved data for physiological conditions should be taken into account, *e.g.* the data on effects of a purified gene product administered at non-physiologically high concentrations are not necessarily relevant for the effect of the gene product expressed from a viral vector under relatively normal physiological conditions. As an example, a search of the Medline literature database for 'interleukin* and toxi*' yields more than 5000 hits, but further evaluation quickly shows that the more meaningful hits describe observed toxicity of interleukins administered as purified protein, at unphysiologically high concentrations.

It should also be taken into consideration that databases like OMIM include information only on those genes for which functions have been shown empirically. This covers only some of the genes that are predicted, *e.g.* from the sequence of the human genome. Several systems have been developed to help manage and display genomic sequences and their annotation, *e.g.* the Ensembl web site[3] (Hubbard et al., 2005, Stalker et al., 2004) and JIGSAW (Allen and Salzberg, 2005), but really reliable prediction is not achieved by any system. Consequently, when genes predicted by bioinformatics analysis are tested empirically in GMOs to evaluate the function of their products, the risk assessment of these applications would have to rely on bioinformatic data, which in terms of the precautionary approach confers a high degree of scientific uncertainty.

The next step in risk assessment is to decide how likely it is that the potential hazards identified would actually lead to a hazardous situation, that could, based on the precautionary approach, require an increased biosafety level. For instance, the use of gene banks containing large numbers of different genes (*e.g.* complete genomic libraries or complete cDNA libraries) cloned in a viral vector, leading to the production of a PHGP in only a small proportion of the GMOs, is usually not seen as an especially hazardous situation; the biosafety class of activities with such a gene bank is determined by the biosafety characteristics of the vectors alone. Only in

[3] http://www.ensembl.org

the case where the gene bank has already been enriched for genes encoding PHGPs by some screening operation, leading to a high proportion of vectors containing these genes, would the presence of PHGPs be taken into consideration to determine the necessary biosafety level.

The question as to what level of expression of a PHGP would require increased physical containment could be tackled on the basis of familiarity. The CMV promoter has been used to drive the expression of a large number of transgenes, which in no case has led to reports of unexpected deleterious effects. This level of expression could therefore be regarded as safe, and overexpression could be operationally defined as an expression level that is at least one order of magnitude higher than the expression of one gene copy under the regime of the CMV promoter.

Finally, the decision as to whether the expression of a PHGP may lead to adverse effects will also depend on the duration of the expression. Use of an adenoviral vector will lead to potentially very high though temporary expression, while lentiviral vectors may lead to much more prolonged though lower expression levels.

When the identification of a potentially hazardous gene product is performed with a high standard of scientific rigor, it may lead to clear conclusions. But even if it does not lead to conclusions, it will lead to clear indications as to why the conclusions cannot (yet) be reached, and what information is necessary. We would therefore recommend that the second stage of hazard identification described above is indeed executed in all cases where indications for a PHGP are found, and that collaborative efforts are made by regulators and researchers to set criteria for the standards of scientific rigor, and for discriminating between 'need-to-know' and 'nice-to-know' research questions.

ACKNOWLEDGEMENTS

This study was commissioned by the Netherlands Ministry of Housing, Spatial Planning and the Environment. Imke Haenen, Caroline Le Poole, Darja Stanič Racman, Martin Batic, and Atanas Atanassov are gratefully acknowledged for their contributions to the discussions.

Received May 18, 2007; accepted December 21, 2007.

REFERENCES

ACGM (2000) Compendium of Guidance from the UK Health and Safety Commission's Advisory Committee on Genetic Modification. Health & Safety Executive, London

Allen JE, Salzberg SL (2005) JIGSAW: integration of multiple sources of evidence for gene prediction. *Bioinformatics* **21**: 3596–3603

Barreto V, Marques R, Demengeot J (2001) Early death and severe lymphopenia caused by ubiquitous expression of the Rag1 and Rag2 genes in mice. *Eur. J. Immunol.* **31**: 3763–3772

Berg P, Baltimore D, Brenner S, Roblin III RO, Singer XF (1975) Summary statement of the Asilomar conference on recombinant DNA molecules. *Proc. Natl. Acad. Sci. USA* **72**: 1981–1984

Dull T, Zufferey R, Kelly M, Mandel RJ, Nguyen M, Trono D, Naldini L (1998) A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72**: 8463–8471

EU (1998) Council Directive 98/81/EC of 26 October 1998, amending Directive 90/219/EEC on the contained use of genetically modified micro-organisms. *Off. J. Eur. Comm.* 05.12.1998: L 330/13

EU (2000) Commission Decision of 27 September 2000 concerning the guidance notes for risk assessment outlined in Annex III of Directive 90/219/EEC an the contained use of genetically modified micro-organisms, 2000/608/EC. *Off. J. Eur. Comm.* 12.10.2000: L 258/43

EU (2001) Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. *Off. J. Eur. Comm.* 17.4.2001: L 106/1

EU (2002) Commission decision of 24 July 2002 establishing guidance notes supplementing Annex II to Directive 2001/18/EC of the European Parliament and of the Council on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC, 2002/623/EC. *Off. J. Eur. Comm.* 30.7.2002: L 200/22

Fallaux FJ, Bout A, Van der Velde I, Van den Wollenberg DJ, Hehir KM, Keegan J, Auger C, Cramer SJ, Van Ormondt H, Van der Eb AJ, Valerio D, Hoeben RC (1998) New helper cells and matched early region 1-deleted adenovirus vectors prevent generation of replication-competent adenoviruses. *Hum. Gene Ther.* **9**: 1909–1917

Fry BG (2005) From genome to "venome": Molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res.* **15**: 403–420

Han JS, Szak ST, Boeke JD (2004) Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **249**: 268–274

HSE (2002) Interim guidance on risk assessment of genetically modified micro-organisms eliciting altered immune responses. HSE Newsletter 31, Health & Safety Executive, London

Hubbard T, Andrews D, Caccamo M, Cameron G, Chen Y, Clamp M, Clarke L, Coates G, Cox T, Cunningham

F, Curwen V, Cutts T, Down T, Durbin R, Fernandez-Suarez XM, Gilbert J, Hammond M, Herrero J, Hotz H, Howe K, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Keenan S, Kokocinsci F, London D, Longden I, McVicker G, Melsopp C, Meidl P, Potter S, Proctor G, Rae M, Rios D, Schuster M, Searle S, Severin J, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Trevanion S, Ureta-Vidal A, Vogel J, White S, Woodwark C, Birney E (2005) Ensembl 2005. *Nucl. Acids Res.* **33**: D447–D453

Jackson RJ, Ramsay AJ, Christensen CD, Beaton S, Hall DF, Ramshaw IA (2001) Expression of Mouse Interleukin-4 by a Recombinant Ectromelia Virus Suppresses Cytolytic Lymphocyte Responses and Overcomes Genetic Resistance to Mousepox. *J. Virol.* **75**: 1205–1210

Mazière P, Enright AJ (2007) Prediction of microRNA targets. *Drug Discov. Today* **12**: 452–458

NIH (2002) NIH Guidelines for research involving recombinant DNA molecules. National Institutes of Health, Bethesda

OECD (1986) Recombinant DNA safety considerations. OECD, Paris

Stalker J, Gibbins B, Meidl P, Smith J, Spooner W, Hotz H-R, Cox AV (2004) The Ensembl Web Site: Mechanics of a Genome Browser. *Genome Res.* **14**: 951–955

Tanino, M, Debily M-A, Tamura T, Hishiki T, Ogasawara O, Murakawa K, Kawamoto S, Itoh K, Watanabe S, de Souza SJ, Imbeaud S, Graudens E, Eveno E, Hilton P, Sudo Y, Kelso J, Ikeo K, Imanishi T, Gojobori T, Auffray C, Hide W, Okubo K (2005) The Human Anatomic Gene Expression Library (H-ANGEL), the H-Inv integrative display of human gene expression across disparate technologies and platforms. *Nucl. Acids Res.* **33**: D567–D572

Yoo S, McKee BD (2004) Overexpression of Drosophila Rad51 protein (DmRad51) disrupts cell cycle progression and leads to apoptosis. *Chromosoma* **113**: 92–101