

RIVM report 601200008/2007

**The effect of REACH implementation on genotoxicity and carcinogenicity testing**

J. van Benthem

Contact: J. van Benthem  
Laboratory for Health Protection Research (GBO)  
E-mail: [jan.van.benthem@rivm.nl](mailto:jan.van.benthem@rivm.nl)

This investigation has been performed by order and for the account of the Ministry of Housing, Spatial Planning and the Environment of the Netherlands (VROM), Directorate General for Environment Protection (DGM), Directorate for Chemicals, Waste and Radiation Protection (SAS), within the framework of project M/601200, Risk Assessment Tools.

RIVM, P.O.Box 1, 3720 BA Bilthoven, the Netherlands; telephone +31 30 274 2021

## Abstract

### **The effect of REACH implementation on genotoxicity and carcinogenicity testing**

Results of an RIVM investigation documented in this report on the effect of REACH, the new European policy on chemicals, show that it is possible to test chemicals for carcinogenicity in an efficient manner without endangering the safety of the product use.

REACH, the Registration, Evaluation and Authorization of Chemicals, is a new EU policy aimed at charting the risks of chemicals to which people are exposed on a daily basis. We are concerned here not only with the approximately 100,000 existing chemicals, but also with new chemicals on the market which are up for testing. The objective of REACH is to realize a more rapid and less expensive method to identify risks, while sparing laboratory animals from tests.

RIVM's aim was to investigate the possibility of scrapping the test for carcinogenicity by employing a more efficient means of testing. This is possible if other tests provide the same data. Examples are (sub)-chronic toxicity or genotoxicity tests measuring DNA damage. Another approach is the Toxicological Threshold Concept (TTC), which yields a fixed value for maximum permissible dose, where the risk is less than one in a million (given as negligible risk).

Exclusive use of *in vitro* (use of cultured cells) tests does not adequately guarantee that DNA damage will not occur. Therefore it is recommended to compare *in vitro* results with results from *in vivo* (use of living organism) experiments. Laboratory animals will still be necessary for measuring DNA damage, but their use will be kept to a minimum through careful choice of testing methods.

Key words: REACH, genotoxicity, carcinogenicity, guidance, test strategie

## Rapport in het kort

### De gevolgen van REACH voor het bepalen van de genotoxiciteit en carcinogeniteit van stoffen

Met het nieuwe Europese stoffenbeleid REACH blijft het mogelijk om stoffen efficiënt te testen op kankerverwekkende eigenschappen zonder dat dit ten koste hoeft te gaan van de veiligheid. Dit blijkt uit onderzoek van het RIVM beschreven in dit rapport.

REACH staat voor Registratie, Evaluatie en Autorisatie van chemische stoffen. Het is een nieuwe regeling om de risico's in kaart te brengen van stoffen waaraan mensen bijna dagelijks worden blootgesteld. Het gaat niet alleen om circa 100.000 bestaande stoffen maar ook om nieuwe stoffen op de markt die moeten worden getest. Doel van REACH is om goedkoper en sneller de risico's te leren kennen en proefdieren te sparen.

Het RIVM heeft onderzocht of een efficiëntieslag mogelijk is door de klassieke test voor carcinogeniteit te schrappen. Dit kan als andere testen de benodigde gegevens leveren. Voorbeelden van andere testen zijn (sub)chronische toxiciteitstesten of genotoxiciteitstesten die de DNA-schade meten die de stoffen veroorzaken. Een andere benadering is het Toxicological Threshold Concept (TTC), dat een vaste waarde aangeeft voor de maximaal toelaatbare blootstelling waarbij het risico minder is dan een op de miljoen (aangeduid als verwaarloosbaar risico).

Uitsluitend testen met gekweekte cellen (*in vitro*) geeft onvoldoende garantie of de stof in een levend organisme (*in vivo*) geen DNA-schade veroorzaakt. Om dit beter te kunnen beoordelen moeten *in vitro* resultaten worden vergeleken met die uit *in vivo* experimenten. Dus blijven proefdieren nodig om DNA-schade te meten, maar wel moet het gebruik van proefdieren tot een minimum worden beperkt door doordacht testen te kiezen.

Trefwoorden: REACH, genotoxiciteit, carcinogeniteit, leidraad, test strategie



## Contents

Summary .....	7
1. Background .....	9
2. REACH requirements to detect the genotoxic and carcinogenic potential of chemicals.....	11
2.1. Requirements for risk characterization .....	11
2.2. (Concept) guidance for genotoxicity and carcinogenicity testing .....	14
2.3. Effects on classification and labelling .....	16
2.4. Effects on risk assessment .....	17
3. Screening of the performance of genotoxicity tests.....	19
4. <i>In vitro</i> genotoxicity tests.....	23
5. <i>In vivo</i> genotoxicity tests .....	27
6. Germ cell genotoxicity testing .....	31
7. Carcinogenicity testing .....	33
8. Integrated (or Intelligent) Test Systems (ITS).....	35
8.1. Prospects of (Quantitative) Structural-Activity Relationships .....	35
8.2. Integration of genotoxicity tests.....	37
9. Conclusion .....	39
10. Recommendations for further research.....	43
Acknowledgements.....	45
References.....	47



## Summary

The aim of the REACH (Registration, Evaluation and Authorization of CHemicals) Regulation is a safety assessment for the existing chemicals which are in use on a daily basis. The requirements under REACH for genotoxicity and carcinogenicity testing have been described and the translation of these requirements has been described into guidance documents (for industry).

The aim of this report is to investigate the effect of REACH implementation on genotoxicity and carcinogenicity testing and to investigate whether at the different levels of testing improvements can be applied.

Implementation of REACH will not hamper classification and labelling for genotoxicity and carcinogenicity as long as this is permitted on scientific data other than exclusively those of germ cell genotoxicity tests or a classical carcinogenicity assay, respectively, as is done now.

REACH may be incomplete as risk assessment for genotoxic carcinogens which is done on the results of the carcinogenicity study which under REACH is only occasionally performed. In the absence of a carcinogenicity assay it is unclear how to derive a dose descriptor for risk assessment of genotoxic carcinogens. Solutions may be found in the minimal toxic dose of (sub)-chronic toxicity studies, the TTC concept or the lowest observed effect dose of *in vivo* genotoxicity tests.

The report demonstrates that it is not possible to determine genotoxicity exclusively on *in vitro* genotoxicity tests due to the high number of false negatives and false positive results. To reduce these false results, it is recommended that a mechanistic justification of both the positive and negative results in genotoxicity tests should be made compulsory.

Only an indication of the actual contribution of the *in vivo* tests within the strategy to detect the genotoxicity of substances could be described. An extensive comparison of the results of *in vitro* and *in vivo* tests covering the same genotoxic endpoint and the same mechanism may elucidate the contribution of *in vivo* tests. Such an evaluation is not only highly recommended since at the end it may justify to evaluate the genotoxic potency of substances exclusively with *in vitro* genotoxicity tests but also because in the absence of a carcinogenicity study, risk assessment for genotoxic carcinogens may be done on genotoxicity data. Special attention should be given to the development of tests which allow the detection of non-genotoxic carcinogens. Since (sub-)chronic and carcinogenicity studies, which are predominantly responsible for the detection of non-genotoxic compounds, are only very occasionally performed, these substances go undetected very easily.

The report also indicates the usefulness of integrated (or intelligent) test systems (ITS) to accelerate evaluations of chemicals and reduce the number of tests. Special attention was given to the *in silico* approach (Q)SARs and integration of tests.

Finally, a number of recommendations or conclusions were given. To determine the intrinsic genotoxic properties of chemicals a battery of three genotoxicity tests appears preferable but two genotoxicity tests are already sufficient. *In vivo* tests are only then allowed if it is demonstrated that the chemical under investigation reaches the target cells and that the *in vivo* follow up test detects the endpoint that showed positive results *in vitro*. The Comet and transgenic mouse assay are good alternatives for the classical *in vivo* tests. Germ cell testing can be reduced to a minimum.

In conclusion, the requirements of REACH and the derived concept guidance for the industry (RIP 3.3-2, concept of February 2007) are adequate. This was to be expected since it uses the knowledge which is acquired with the current almost identical guidance (TGD, 2003). However, because of this similarity it is questionable whether under REACH the costs, number of tests and above all the number of laboratory animals for genotoxicity and carcinogenicity testing will actually be reduced. To reach this latter aim, improvements may be necessary which, fortunately, can be enforced without giving up in essence much on safety assessment. For instance, waiving of further testing for compounds with minimal human exposure that are exclusively weakly positive in the gene mutation test in bacteria or of the carcinogenicity assay for all *in vivo* genotoxic compounds.



# 1. Background

About 30,000 existing chemicals are in use by humans on a regular basis. Many of these chemicals are high production volume chemicals. At the moment only a small portion ( $\approx 3\%$ ) of these chemicals is completely analyzed according to present requirements.

In the White Paper (EC, 2001) a strategy for a future chemicals policy, was outlined. Following the principles and conclusions of the White Paper a new system for the Registration, Evaluation and Authorization of Chemicals (REACH) was introduced (EU, 2003). The aim of the REACH Regulation is a safety assessment for the remaining 97% of the existing chemicals by describing the future chemical legislative system for the EU.

The REACH regulation requires toxicological information from every chemical with a production volume  $> 1$  tpa. The amount of information required is dependent on the tonnage level: a higher production volume means that more information is needed. Obviously, it is to be expected that in the years after the implementation of REACH large amounts of regulatory toxicity tests have to be carried out to fill the data gaps. Predictions of the EU indicate that the greatest number of tests is necessary for skin sensitization (35 % of all chemicals), eye irritation (24 % of all chemicals) and mutagenicity (22 % of all chemicals) (Pedersen *et al.*, 2003). The animal need is greatest for reproduction toxicology (two generation and developmental studies), mutagenicity and skin sensitization (Van der Jagt *et al.*, 2004).

Since REACH involves many chemicals, many tests, many laboratory animals and high costs, REACH demands due to logistic, economical and ethical reasons new tests and strategies. Alternative test systems that will require less time, animals and money and have better or at least comparable accuracy compared to the accuracy of classical tests will be essential. This also includes the need for development of reliable non-testing methods like read across, category approach and *in silico* methods like QSARs. Nevertheless, classification and labelling as well as risk assessment have to be as reliable as before.

The aim of the present report is to investigate the effect of REACH implementation on genotoxicity and carcinogenicity testing and to list whether at the different levels of testing, adaptations and/or improvements can be applied. The requirements for genotoxicity testing and carcinogenicity testing have been described (EU, 2006). With the REACH Implementation Program (RIP 3.3-2, concept of February 2007), these requirements are translated into a (draft) guidance (for industry). In chapter 2 the requirements, the concept guidance, and the effects of these on classification and labelling and on risk assessment will be discussed. In the following chapters the specific levels of the strategy come up for discussion. Since in these chapters

databases will be used to demonstrate the performance of the various test systems these databases are discussed in chapter 3. The discussion on the *in vitro* tests (chapter 4) focus on the question how many tests are preferable. Also the possibility to determine the genotoxic potency of compounds with only *in vitro* tests will be considered. In chapter 5 the same is done for the *in vivo* tests with special attention on the question to what extend these *in vivo* tests have added to conclusions already based on *in vitro* tests. The idea to waive germ cell mutagenicity tests and the carcinogenicity test is put forward in chapters 6 and 7, respectively. Finally, in chapter 8 alternative approaches will be discussed.

## 2. REACH requirements to detect the genotoxic and carcinogenic potential of chemicals

The aim of testing for genotoxicity and/or carcinogenicity is to assess the potential of chemicals to cause heritable damage to humans or to be genotoxic carcinogens. The information requirements for mutagenicity and carcinogenicity are described in REACH Annexes VI to XI, which specify the information that has to be submitted for registration and evaluation purposes. The information is used for risk characterization and for classification and labelling of chemicals.

The information is required for substances produced or imported in quantities of >1 tonne per annum (tpa) and is strongly tonnage triggered. When a higher tonnage level is reached, the requirements of the corresponding Annex have to be settled. Pre-existing toxicity data, information about the intended use of the substance and human exposure to the substance will influence the information requirements. The REACH Annexes shall thus be considered as a whole, and in conjunction with the overall requirements of registration, evaluation and the duty of care.

### 2.1. Requirements for risk characterization

The standard information required for all substances manufactured or imported in quantities of 1 tpa or more (Annex VII) are data from an *in vitro* gene mutation study in bacteria (Ames test). Further mutagenicity studies shall be considered in case of a positive result. Although not clearly described these studies should be linked to the *in vitro* studies of Annex VIII.

For all substances manufactured or imported in quantities of 10 tpa or more (Annex VIII) the standard information required additional to that from annex VII are data from an *in vitro* cytogenicity study in mammalian cells. In case of a negative result in the *in vitro* cytogenicity study in mammalian cells and in the *in vitro* gene mutation study in bacteria, data from an *in vitro* gene mutation study in mammalian cells are required. Appropriate *in vivo* mutagenicity studies shall be considered in case of a positive result in any of the genotoxicity studies.

If there are no results available from an *in vivo* study, for substances manufactured or imported in quantities of 100 tpa or more (Annex IX) or 1000 tpa or more (Annex X), additional to the requirements of annexes VII and VIII, the registrant has to propose an appropriate *in vivo* somatic cell genotoxicity study. Depending on the quality and relevance of all the available data, a second *in vivo* somatic cell test may be necessary.

For substances that were positive in cytogenetic tests, it has to be established whether they specifically act as aneugens, if not known already. This should be done using *in vitro* methods and may be helpful in risk evaluation

If there is a positive result from any *in vivo* somatic cell study available, the potential for germ cell mutagenicity should be considered on the basis of all available data, including toxicokinetic evidence. If no clear conclusion about germ cell mutagenicity can be made, additional investigations shall be considered.

REACH requires that carcinogenicity is considered at all tonnage levels taking into account information from all available relevant sources. The minimum information for carcinogenicity at Annexes VII, VIII and IX is equivalent to that required for the mutagenicity endpoint. As such, this will not lead to classification of a substance as a carcinogen, but this evidence should be taken into account in risk assessment. Substances shown to be *in vivo* mutagens should be assumed to be potentially carcinogenic. Furthermore, the data from other toxicity tests (*e.g.* repeated dose toxicity or reproduction toxicity tests) may be informative about a possible carcinogenic potential and on potential mode(s) of action underlying the carcinogenic effect.

The actual performance of a carcinogenicity test is only specified at Annex X if the chemical is considered mutagenic class 3 or there is evidence from the repeated dose study(ies) that the substance is able to induce hyperplasia or pre-neoplastic lesions and the substance has a widespread dispersive use or there is evidence of frequent or long-term human exposure. Due to economical and ethical disadvantages of the classical carcinogenicity assay, REACH preferably requires that the putative carcinogenic potential of a test substance is decided without the conduct of such tests. Moreover, if the substance is classified as mutagenic class 1 or 2, the default presumption would be that a genotoxic mechanism for carcinogenicity is likely and a carcinogenicity test will normally not be required.

At all levels the REACH program demands that, before (additional) tests are carried out, all available *in vitro* data, *in vivo* data, historical human data, data from validated (Q)SARs and data from structurally related substances (read-across approach) shall be assessed first. This may lead to waiving of unnecessary tests. Moreover, if a registrant wishes to undertake any test for substances at Annexes IX or X that requires the use of laboratory animals, then there is a need to make a proposal to the European Chemicals Agency first. Tests can not be performed without agreement of the Agency.

Chemicals with a production volume < 1 tpa, for which the REACH program does not have testing requirements, have to be registered and all available information has to be submitted. This information will probably predominantly contain non-testing data.

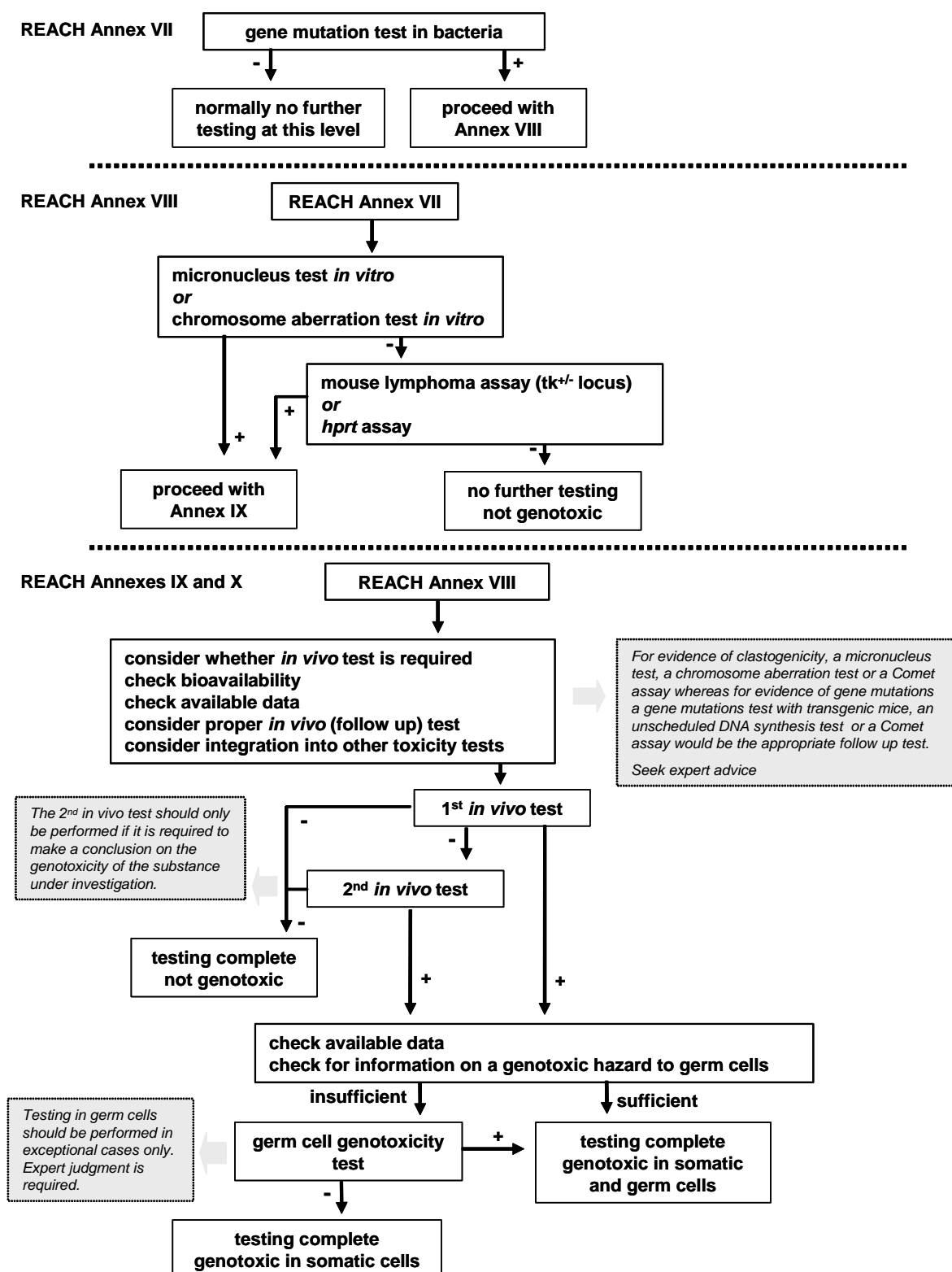


Figure 1: Proposed flow chart for mutagenicity testing under REACH. (RIP 3.3-2 EWG 7 Report on Mutagenicity, fourth draft, February 2, 2007)

In conclusion, these requirements are adequate and certainly not excessive seen the experience obtained from the current almost identical requirements. They guarantee a reliable determination of the genotoxic and carcinogenic potency of compounds.

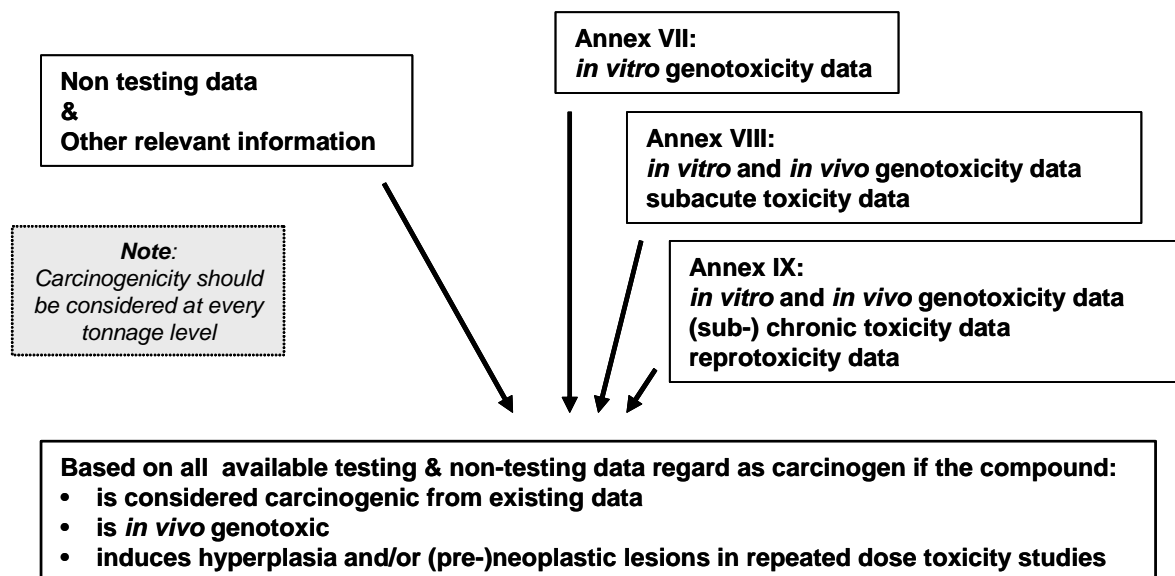
Accepting that an aim of REACH is to reduce the costs, the number of tests and above all the number of laboratory animals, still improvements can be extracted without giving up in essence much on safety assessment. For instance for compounds with minimal human exposure and which are weakly positive in genotoxicity tests, further testing may be superfluous. Furthermore, since it is generally accepted the default should be that *in vivo* mutagens can be considered as genotoxic carcinogens unless there is scientific evidence to the contrary. Hence, a carcinogenicity assay can be waived for all genotoxic carcinogens and not only for those classified as class 2 or 1 mutagens.

## 2.2. (Concept) guidance for genotoxicity and carcinogenicity testing

In the REACH Implementation Program (RIP 3.3-2, concept of February 2007) guidance documents are formulated to guide the industry through the requirements. These guidance documents translate the requirements of REACH into a rather flexible strategy by mentioning that in some specific circumstances tests can be waived. In the fourth drafts (February 2007; figures 1 and 2) of the concept guidance for genotoxicity and carcinogenicity, the main possibilities to waive tests are given below provided that a sound scientific justification has been given.

- Tests need not be performed if it is not technically possible to do so, or if they are not considered necessary in the light of current scientific knowledge.
- If at Annex VII the bacterial gene mutation test is positive, it is important to consider the possibility of the substance being genotoxic in mammalian cells. The need for further testing according to Annex VIII depends on an evaluation of all the available information relating to the genotoxicity of the substance
- At Annex VIII, when both the mammalian cell tests are negative but there was a positive result in the bacterial test, it will be necessary to decide whether any further testing is needed on a case-by-case basis.
- The second *in vitro* genotoxicity test in mammalian cells (gene mutation test) may be omitted if it can be demonstrated that this mammalian cell test will not provide any further useful information about the potential *in vivo* mutagenicity of a substance. This should be evaluated on a case-by-case basis.
- The second *in vivo* test should only then be performed if the *in vitro* data show that the substance has the potential to induce both gene and chromosome mutations and the first *in vivo* test is negative but has not addressed both genotoxic endpoints comprehensively.

Annexes VII - IX



Annex X

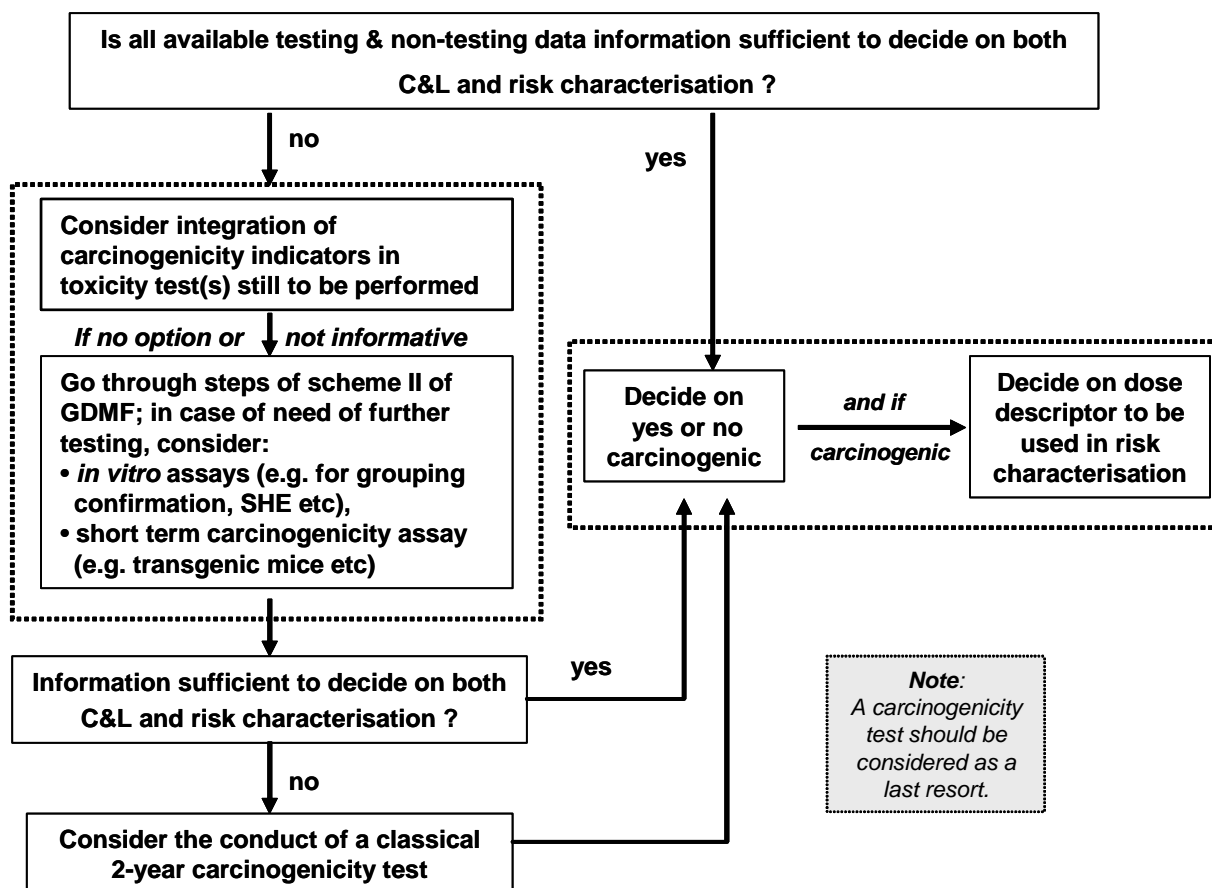


Figure 2: Proposed flow chart for carcinogenicity testing under REACH. (RIP 3.3-2 EWG 7 Report on Mutagenicity, fourth draft, February 2, 2007)

- For substances classified as class 3 mutagen (positive in *in vivo* somatic cell genotoxicity tests), it should first be evaluated whether classification in class 2 for mutagenicity is possible which allows the waiving of the carcinogenicity test. If an upgrade of classification is not possible, these substances will be regarded as genotoxic carcinogens, unless there is clear evidence to indicate the contrary. On a case by case basis it should be decided if and how the carcinogenic potential has to be explored. The possibility to derive within a reasonably small range a reliable dose discriminator on other tests than a carcinogenicity test can be a justified reason to waive a carcinogenicity assay.
- Only occasionally a carcinogenicity test may be justified if there are clear suspicions that the substance may be carcinogenic, and available information is not conclusive in this, both in terms of hazard and potency. However, a carcinogenicity study should be considered only as a last resort

It can be concluded that the proposed concept guidance described above (RIP 3.3-2, concept of February 2007) is adequate. It uses the knowledge which is acquired with the current almost identical guidance (TGD, 2003). The choice of the specific tests is understandable. For genotoxicity, exclusively tests which directly measure permanent DNA damage are recommended. The only exception is the *in vivo* unscheduled DNA synthesis test (indicator test) because a well validated alternative is not (yet) generally accepted.

### 2.3. Effects on classification and labelling

Although REACH is predominantly developed for safety assessment, it has an effect on classification and labelling. The rules for classification and labelling of genotoxic and carcinogenic compounds are described in EU context (EEC, 1967). Council Directive 67/548/EEC will be brought in coherence with the ‘Global Harmonized System of Classification and Labelling of chemicals’ (GHS, 2006).

For genotoxicity, classification and labelling is based on the results of germ cell genotoxicity tests which are under REACH only occasionally performed. Positive results in *in vivo* genotoxicity tests may automatically lead to classification as class 3 mutagen. If on the basis of expert judgment it is concluded that there is sufficient reliable information to consider the substance to pose a mutagenic hazard to germ cells, further germ cell testing is not essential; the chemical is immediately classified as a class 2 mutagen. Since at all levels the REACH program demands that all available data including epidemiological data, which are not specifically included in the REACH strategy, shall be assessed first, also the class 1 mutagens will be recognised.



Classification and labelling for carcinogenicity depends on the outcome of the carcinogenicity bioassay which under REACH is only occasionally performed. If there is evidence from (semi-)chronic toxicity, genotoxicity tests (mutagen class 1 or 2) or even non-testing data that a chemical poses a carcinogenic risk and on the basis of expert judgment the classical carcinogenesis assay is waived, the chemical should be classified for carcinogenicity. If this evidence is lacking a carcinogenicity assay may be essential after all. An *in vitro* transformation assay or a short term carcinogenicity assay with transgenic animals may then be considered rather than a two year bioassay. However, *in vitro* data only give an indication to a putative carcinogenic potential whereas the assay with transgenic mice may be suitable for hazard identification. The suitability for risk assessment of the latter test is currently under discussion.

## 2.4. Effects on risk assessment

Under REACH the classical carcinogenicity assay will only be occasionally performed. Although this looks like a benefit of the REACH program, it hampers risk assessment for genotoxic carcinogens which relies on these carcinogenicity studies. After recognized as a genotoxic carcinogen, a dose descriptor has to be derived, that allows the derivation of the so-called Derived-Minimal-Effect-Level (DMEL) that, in turn, is used to establish exposure levels of minimal concern. In the absence of a carcinogenicity assay it is unclear how to derive a dose descriptor for risk assessment of genotoxic carcinogens. A possibility is to identify the minimal toxic dose in (sub)-chronic toxicity studies and applying a large assessment factor. The Threshold of Toxicological Concern (TTC) concept may give relief (Kroes *et al.*, 2004, 2005). The TTC concept identifies a *de minimis* exposure value for all chemicals, including genotoxic carcinogens, below which there is no appreciable risk to human health for any chemical. Although routinely performed genotoxicity tests are not designed for risk assessment, the Lowest Observed Effect Dose (LOED) of *in vivo* genotoxicity tests may be an alternative tool in risk assessment (see TGD, 2003).

Clearly for the derivation of the descriptor dose expert judgment is indispensable.



### 3. Screening of the performance of genotoxicity tests

In the following paragraphs the impact of REACH on the different levels of the strategy will be discussed. Particularly performance (~ predictive value) data will be used. To determine the performance of these tests it is essential to consider to which endpoint the performance should be opposed to. The endpoints of genotoxicity testing, gene mutations and chromosomal aberrations are 'genotoxic' endpoints. They are just 'stopovers' on the way to the 'real' endpoint which is the disease resulting from the mutations. Cancer is a disease of somatic cells which is strongly linked to the occurrence of mutations. Consequently, the performance of the genotoxicity tests can

*Table 1: Terms used to describe the performance of the genotoxicity tests.*

	carcinogens	non-carcinogens
genotoxicity positive	A	B
genotoxicity negative	C	D

sensitivity	% correct predicted carcinogens	$A * 100 / A + C$
specificity	% correct predicted non-carcinogens	$D * 100 / B + D$
concordance	% correctly predicted carcinogens and non-carcinogens	$(A + D) * 100 / A + B + C + D$
pos. predictivity	% correctly predicted carcinogens among positive results	$A * 100 / A + B$
neg. predictivity	% correctly predicted non-carcinogens among negative results	$D * 100 / C + D$

be determined by assessing the predictivity for cancer. It does, however, not mean that these tests show the same performance for genetic heredity.

Table 1 describes the terms used. Next to the definition of Table 1, sensitivity and negative predictivity give also an indication of the number of false negative results (negative results in genotoxicity test obtained with carcinogens); specificity and positive predictivity for false positive results (positive results in genotoxicity test obtained with non-carcinogens). Although frequently used in the present report because they are generally accepted, in fact, false positive and false negative results

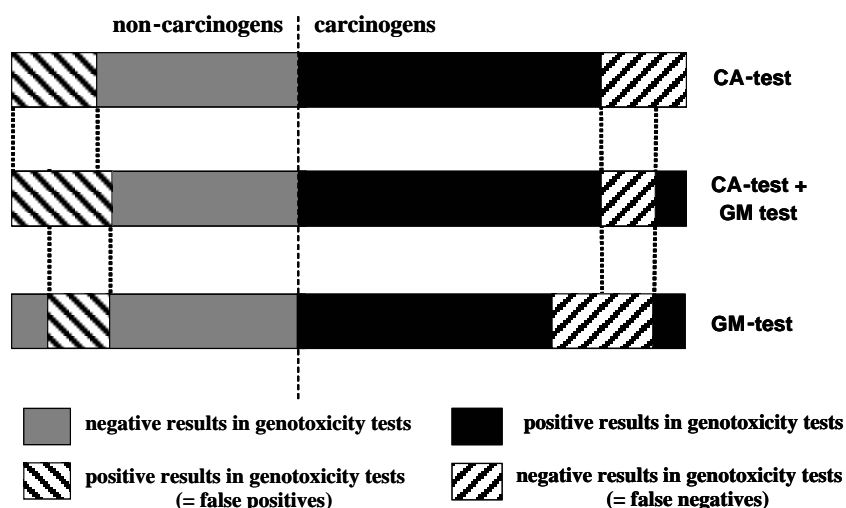


Figure 3: Combinations of tests results in an increase of false positives but a decrease in false negatives.

do not exist! These results are correct results in that specific test. False negative results are better described as unexpected negative results obtained with carcinogens and likewise false positive results as unexpected positive results with non-carcinogens.

When taking cancer as the endpoint, one has to take into consideration that cancer can arise along genotoxic and non-genotoxic mechanisms. In the evaluations on the performance of tests this difference between genotoxic and non-genotoxic carcinogens is hardly ever made. Carcinogens with a non-genotoxic mode of action will always score negative and are easily considered false negatives in genotoxicity tests whereas in fact they are 'correct' negatives. On the basis of the present report it is recommended to assess the predictive performance of genotoxicity tests when the mode of action is taken into account.

For each of the two genotoxic endpoints (chromosome aberrations and gene mutations) different genotoxicity tests exist. This also has an impact on the number of false negatives. If a carcinogen scores negative in a specific genotoxicity test it (i) mostly is the result from a genotoxicity test that does not cover the genotoxic endpoint which makes the substance tested a carcinogen rather than that it indicates to (ii) a non-genotoxic carcinogen or (iii) to a real false negative result. For instance, a carcinogen which predominantly induces chromosome aberrations will generally score positive in a chromosome aberration test but (correctly) negative in gene mutation tests.

The value of the evaluation on the performance of genotoxicity tests in relation to their predictivity for carcinogenicity is strongly dependent on the databases used. The quality of the tests and the conclusions drawn from the tests contribute to the reliability of the database. Therefore, evaluators often re-evaluate older tests or tests with an apparent lower quality. Obviously, the total number of chemicals in the

database and particularly the number of non-carcinogens is important. Most databases suffer from a very low number of non-carcinogens which strongly influences specificity, negative predictivity but also concordance which generally is used to describe the performance of tests. The databases used for the results discussed in this report contain both genotoxic and non-genotoxic carcinogens and do not distinguish between rodent carcinogens and human carcinogens by which nothing can be said about the predictivity of the genotoxicity tests for human cancer risk.

Finally, since every individual test suffers from false negatives and false positives, in combinations of tests the number of false negatives will decrease (Figure 3) because if one test in a combination of tests scores positive, the worse case approach is followed and the chemical is considered positive (decreasing the number of false negatives). Contrary, a chemical is only then negative if all tests performed score negative. The number of false positives, consequently, will increase in combinations.



## 4. *In vitro* genotoxicity tests

The major role of the *in vitro* tests is to assess the genotoxic potential of chemicals as fast and reliable as possible but with as little as possible tests and costs. The aim of this chapter is to discuss the number of tests necessary to guarantee a reliable determination of the intrinsic genotoxic potency of compounds and to investigate whether it may be justified to determine the genotoxic potential of chemicals exclusively on *in vitro* tests.

For the two endpoints of concern, gene mutations and structural and numerical chromosomal aberrations, well established tests with OECD guidelines either available or in preparation, can be performed.

The gene mutation endpoint is covered by the gene mutation tests in bacteria (Ames test; OECD 471; Ames *et al.*, 1975; Maron and Ames, 1983) and the *in vitro* mammalian cell gene mutation test (OECD 476; Moore *et al.*, 1987; Liber and Tilly, 1982). An advantage of the mouse lymphoma assay using the *tk* gene is that both gene mutations (large colonies) and chromosomal aberrations (small colonies, due to growth retardation) can be recognized.

Structural chromosome aberrations can be detected with the *in vitro* chromosome aberration test (OECD 473, Evans, 1976) and both structural and numerical chromosome aberrations with the *in vitro* micronucleus test (OECD 487 in preparation, Fenech and Morley, 1985).

Although many papers have been published on this subject, to demonstrate the performance of the *in vitro* tests the most recent evaluation on *in vitro* tests (Kirkland *et al.*, 2005), screening the most popular *in vitro* genotoxicity tests for their ability to discriminate between carcinogens and non-carcinogens, was used. This evaluation was performed on 756 chemicals considered as rodent carcinogens (based on findings in at least one sex rats or mice) and 183 non-carcinogens. Many genotoxicity tests on these chemicals were re-evaluated by experts since interpretation of data changed over the years. Although the database used by Kirkland *et al.* (2005) is very extensive, it still suffers from the shortcomings described in chapter 3; particularly the distribution between carcinogens and non-carcinogens is not balanced.

In Table 2 the performance of the individual tests is shown. The concordance (between 60 and 65%) shows that 2 out of 3 chemicals are correctly predicted. However, it also indicates to both false negatives and false positives. This may not be as bad as it looks because, as mentioned earlier, a number of carcinogens will score 'correctly' negative in the individual genotoxicity tests because the genotoxic endpoint which is not the reason for carcinogenicity of the chemical, is measured. Since two genotoxic endpoints, gene mutations and chromosomal aberrations, have to

be investigated, it is more meaningful to evaluate the performance of combinations of tests covering both genotoxic endpoints.

Table 2: Performance of the individual *in vitro* tests in detecting rodent carcinogens and non-carcinogens (data from Kirkland *et al.*, 2005)

	Ames	MLA	MN <sub>vitro</sub>	CA <sub>vitro</sub>
sensitivity	58.8	73.1	78.7	65.6
specificity	73.9	39.0	30.8	44.9
concordance	62.5 (484/717)	62.9 (220/350)	67.8 (78/115)	59.8 (292/488)
pos. predictivity	87.4	73.7	79.5	75.5
neg. predictivity	36.8	38.3	29.6	33.5

Ames: Ames test (*in vitro* gene mutation assay in bacteria); MLA: mouse lymphoma assay; MN<sub>vitro</sub>: *in vitro* micronucleus test; CA<sub>vitro</sub>: *in vitro* chromosome aberration test

The evaluation by Kirkland *et al.* (2005) for combinations of two or three assays (Table 3) demonstrated the expected higher sensitivity (less false negatives) accompanied by a decreased specificity (more false positives) compared to single tests (see also chapter 3 Figure 3). Kirkland *et al.* (2005) demonstrated an increased concordance for combinations of two but not for three tests. The combinations Ames test/chromosome aberrations and mouse lymphoma assay/chromosome aberrations, covering both genotoxic endpoints, appeared sensible combinations. However, combinations with the chromosome aberration test show a lower sensitivity and thus more false negatives than those with the micronucleus test. The latter combinations should be recommended by regulatory bodies. Unfortunately, combinations including the MN-test resulted in dramatic decreases in specificities, probably due to low number of chemicals which were tested negative in all tests of the various combinations. In a retrospective validation study an expert panel (Albertini *et al.*, in prep.) concluded that the *in vitro* micronucleus test can be regarded as sufficiently validated and can be recommended as an alternative for the *in vitro* chromosome aberration test.

A battery of three *in vitro* tests, Ames test/ mouse lymphoma assay (*tk*<sup>+/-</sup> locus)/ chromosome aberration test or micronucleus test is often considered preferable not because it is more accurate as Table 3 shows but because the mammalian mouse lymphoma assay may confirm the results of both the bacterial Ames test and the chromosomal aberration test covering both genotoxic endpoints twice. In the (concept) guidance document for genotoxicity testing (RIP 3.3-2, concept of February 2007), it is suggested that if in the case of a negative Ames test and a negative chromosome aberration tests, a gene mutation test in mammalian cells is required, under special circumstances this test can be waived if it can be scientifically



demonstrated that it will not provide any further useful information. The data from Table 3 support this suggestion. The performance of combinations of the Ames tests

Table 3: Performance of a battery of *in vitro* tests in detecting rodent carcinogens and non-carcinogens (data from Kirkland *et al.*, 2005)

	Ames + MLA <sup>1</sup>	Ames + MN <sub>vitro</sub> <sup>1</sup>	Ames + CA <sub>vitro</sub> <sup>1</sup>	MLA + MN <sub>vitro</sub> <sup>1</sup>	MLA + CA <sub>vitro</sub> <sup>1</sup>	Ames + MLA + MN <sub>vitro</sub> <sup>2</sup>	Ames + MLA + CA <sub>vitro</sub> <sup>2</sup>
sensitivity	89.2	94.4	82.2	89.9	86.1	90.7	84.7
specificity	27.4	4.9	33.1	2.8	21.3	5.0	22.9
concordance	75.5 (423/560)	81.8 (353/433)	71.1 (445/626)	69.4 (213/307)	69.2 (324/468)	67.6 (50/74)	64.8 (193/298)
pos. predictivity	81.2	85.8	80.7	75.1	75.6	72.1	69.8
neg. predictivity	42.0	12.5	35.3	7.7	35.1	16.7	41.5

1: if 1 or 2 tests are performed; 2: if all three tests were performed

Ames: Ames test (*in vitro* gene mutation assay in bacteria); MLA: mouse lymphoma assay; MN<sub>vitro</sub>: *in vitro* micronucleus test; CA<sub>vitro</sub>: *in vitro* chromosome aberration test

and the chromosome aberrations or micronucleus test hardly increases with the introduction of the mouse lymphoma assay into the combination. Moreover, a slight increase of the number of false positives may occur.

Table 3 demonstrates that it seems not sensible to determine the genotoxicity of chemicals exclusively with *in vitro* tests. Combinations of *in vitro* tests are essential to cover both endpoints but give rise to a number of false positive and false negative results.

Since a year and a day a certain level of real false negative chemicals is accepted. Fortunately, Kirkland *et al.* (2005) showed that in their database the mechanism of action for carcinogenicity of 80% of the false negative chemicals is known to be non-genotoxic. Yet the remaining 20% may still be carcinogenic. REACH will not give insight in these putative carcinogens. On the other hand the false positives are a bigger problem for industry than for regulators. Promising chemicals which score false positive in genotoxicity tests, will be skipped from further development. On the other hand, false positive results in *in vitro* genotoxicity test may trigger many unnecessary *in vivo* tests using large numbers of laboratory animals.

Of course more accurate *in vitro* tests could certainly significantly reduce the number of false positive and false negative results. On the other hand, if for every positive or negative result in a genotoxicity test a mechanistic justification is demanded this will not only lead to more appropriate follow up tests but will also allow recognition of false positive or false negative results. Apparently, genotoxicity testing suffers from the lack of such a mechanistic justification. At least a study to the mechanisms behind

(false) negative and (false) positive results should be recommended. Under REACH knowledge on these mechanisms is all the more important because carcinogenicity tests are only occasionally available.

Recently, in a workshop organized by ECVAM (Ispra, Italy) the high rate of false negative and particularly false positive results in genotoxicity tests was addressed (Kirkland *et al.*, 2007). During the workshop it was investigated (i) whether it is possible to choose existing cell systems which give lower rates of false results, (ii) whether modifications of existing guidelines or cell systems may result in lower false (positive) results, and (iii) the performance of new systems showing promise of improved specificity was reviewed.

It was concluded that better guidance on the likely mechanisms (high cytotoxicity, high passage number of cell lines, *p53* status, DNA repair status, etc) resulting in positive results not relevant for humans and on how to obtain evidence for those mechanisms, is needed. The development of new test systems was not of the highest priority but if they will be developed then cell systems of human origin which are *p53* and DNA repair proficient, and have defined phase 1 and phase 2 metabolism, offer the best hope to reduced false positive or negative results in the future. A collaborative research program is needed to identify and evaluate (new) cell systems with appropriate sensitivity but improved specificity.

## 5. *In vivo* genotoxicity tests

The major aim of the *in vivo* genotoxicity tests is to confirm the positive results of *in vitro* genotoxicity tests and to recognize and eliminate the chemicals which score false positive in the *in vitro* tests. In the guidance it is suggested that by default this aim should be reached with only one *in vivo* genotoxicity test. A second *in vivo* test (putative requirement of Annex VIII) is only then justified if the *in vitro* data show the substance to have potential to induce both gene and chromosome mutations and the first *in vivo* test has not addressed this comprehensively. Risk managers should consider that *in vivo* tests should only be allowed for indispensable compounds for which no alternative is available.

The aim of this chapter is to discuss which *in vivo* test is preferable and in how far the *in vivo* tests contribute to the data obtained from *in vitro* tests.

To reduce the number of unnecessary *in vivo* genotoxicity test, before *in vivo* testing is undertaken, it should be demonstrated that the chemical under investigation reaches the target cells of the *in vivo* test, for instance by toxicokinetic, toxicodynamic or distribution studies. If it can not be reasonable expected from all the properties of the test chemical that the specific target tissue of genotoxicity tests will be adequately exposed, further testing with *in vivo* tests is not desirable because it would merely produce irrelevant (negative) results. The compound should be considered as an *in vitro* or intrinsic mutagen. Because it can not be excluded that the compound is a mutagen classification and labelling may be desirable which is not possible under the current guideline for classification and labelling.

Secondly, since there are no criteria for further testing, a case by case approach will have to indicate which test(s) may be best performed. The *in vivo* follow up test needs to be a logic follow up, i.e. the test should cover the same genotoxic endpoint as the one which showed positive results *in vitro*. For instance, if a chemical appeared a clastogen under *in vitro* conditions than further testing is only justified with an *in vivo* test for clastogenicity.

Finally, the classical *in vivo* tests suffer from a certain tissue restriction (bone marrow, peripheral blood cells, hepatocytes). Considering that *in vivo* testing is in fact a prescreen for cancer, it is obvious that the value of the *in vivo* tests dramatically increases if the target tissue(s) for carcinogenicity are investigated. Therefore, tests without obvious tissue restriction should be preferable as follow up test.

An appropriate test without tissue restriction to investigate the occurrence of gene mutations is the *in vivo* gene mutation test with transgenic mice (Gossen *et al.*, 1989; Kohler *et al.*, 1991). Since this test is not yet fully validated and an OECD guideline does not exist, as a surrogate the *in vivo* unscheduled DNA synthesis test (OECD 486;

Table 4: Performance of the individual *in vivo* tests in detecting rodent carcinogens and non-carcinogens

	CA <sub>vivo</sub> <sup>1</sup>	MN <sub>vivo</sub> <sup>1</sup>	UDS <sup>3,4</sup>	Comet <sup>2</sup>	Tg <sup>3</sup>
sensitivity	43.6	36.4	69	78.1	78
specificity	66.7	77.8	60	80.0	69
concordance	51.2 (42/82)	50.0 (41/82)	68 (25/37)	78.4 (149/190)	77 (81/105)
pos. predictivity	72.7	76.9	92	95.4	95
neg. predictivity	36.7	37.5	23	40.7	31

1: Kim and Margolin, 1999; 2 Sasaki *et al.*, 2000; 3: Lambert *et al.*, 2005. 4: Since the number of non-carcinogens in the database is low, firm conclusions on the specificity and negative predictivity for these tests are unreliable.

CA<sub>vivo</sub>: *in vivo* chromosome aberration assay; MN<sub>vivo</sub>: *in vivo* micronucleus test; UDS: unscheduled DNA synthesis test; Comet: comet assay; Tg *in vivo* gene mutation assay with transgenic mice.

Ashby *et al.*, 1985; Butterworth *et al.*, 1987) can be considered. However, the latter test is an indicator test measuring the repair of DNA damage in hepatocytes of rats from which never can be proven that when not repaired it would have lead to a mutation.

The structural chromosomal aberration endpoint is covered by the mammalian bone marrow chromosome aberration test (OECD 475; Adler, 1984; Preston *et al.*, 1987) whereas both numerical and structural chromosome aberrations can be determined with the mammalian erythrocyte micronucleus test (OECD 474; Heddle *et al.*, 1983). The single cell electrophoresis assay or Comet assay (Tice *et al.*, 1990) is a good alternative without tissue restriction. It allows the determination of single or double-strands DNA breaks, phosphotriesters and alkali label sites in every cell type which can be cultured. This test is not yet validated and an OECD guideline is not available.

Table 4 shows the performance of the different *in vivo* assays. These data may suffer from the shortcomings of databases as already mentioned in chapter 3. Particularly, the total number of chemicals in the databases is rather low and the ratio between (genotoxic and non-genotoxic) carcinogens and non-carcinogens is not balanced. The database used by Lambert *et al.* (2005) was build to promote the *in vivo* gene mutation test with transgenic mice and therefore is biased towards chemicals investigated in the transgenic mouse assay.

The UDS test and particularly the Comet assay and the gene mutation test with transgenic animals perform relatively well which is demonstrated by the rather high sensitivity and specificity. The sensitivities of the chromosome aberration test and the

Table 5: *In vivo* confirmation of *in vitro* positives (Lambert *et al.*, unpublished results).

	sensitivity
<i>In vivo</i> CA vs <i>in vitro</i> CA	72 (21/29)
<i>In vivo</i> CA vs <i>in vitro</i> MN	74 (32/43)
<i>In vivo</i> MLA vs <i>in vitro</i> UDS	71 (15/21)
<i>In vivo</i> MLA vs <i>in vitro</i> Tg	75 (28/37)

CA: chromosome aberration assay; MN: micronucleus test; MLA: mouse lymphoma assay; UDS: unscheduled DNA synthesis test; Tg: *in vivo* gene mutation assay with transgenic mice.

micronucleus test are low. However, Lambert *et al.* (2005) reported a much higher sensitivity for these latter tests. The positive predictivity of the Comet, transgenic mice and the unscheduled DNA synthesis test is very high indicating that a positive result points with more than 92% certainty to a genotoxic carcinogen. The carcinogenicity assay in this case may be redundant. Since the Comet assay and the gene mutation assay with transgenic animals are preferable as was argued before, this may be another point to consider them as the first to consider *in vivo* follow up tests. In the chromosome aberration and the micronucleus test the positive predictivity is lower and at the border whether it is sufficient to consider a positive result as an indication for a genotoxic carcinogen.

A main question is what the contribution is of the *in vivo* genotoxicity tests to what is already known from the *in vitro* tests. In other words how often results from *in vivo* tests do change the conclusion based on *in vitro* genotoxicity tests. If this occurs only occasionally one may wonder whether it is worth the effort to perform the animal consuming *in vivo* tests and provided that the mechanistic background of the *in vitro* results is available, it may justify to evaluate the genotoxic potency of substances exclusively with *in vitro* genotoxicity tests.

To determine the contribution of the *in vivo* genotoxicity tests in relation to the conclusion based on *in vitro* tests is rather difficult. First as for the *in vitro* results a mechanistic justification for both the positive and negative results should be available. Next an extensive comparison has to be carried out between the results of *in vitro* and *in vivo* tests covering the same genotoxic endpoint and the same mechanism.

A first attempt can be deduced from table 5 which demonstrates that around 75% of the *in vitro* positive chemicals also score positive in the corresponding *in vivo* test. Apparently, 25% in this comparison were the false positives recognized by the *in vivo* tests. However, the database used is biased as is mentioned before (see chapter 3). Seen the importance of genotoxicity testing for risk assessment, and since under REACH a carcinogenicity study is only occasionally available it may be necessary to carry out risk assessments based only on genotoxicity data. An appropriate evaluation to the contribution of *in vivo* tests in the determination of the genotoxic potency of

compounds and the development of a strategy to allow for inclusion of test results in risk assessment system is highly recommendable.

## 6. Germ cell genotoxicity testing

The moment germ cell genotoxicity assays come up, the chemical under investigation is already considered genotoxic on the basis of *in vivo* genotoxicity tests in somatic cells. However, chemicals that induce mutations in stem cells are of primary importance in germ cell risk assessment because mutations in stem cells persist and can be transmitted during the remainder of the reproductive lifespan of the individual (Allen *et al.*, 1995). Therefore, germ cells tests play an decisive role in classification and labelling of chemicals for genotoxicity and they are still part of the REACH requirements.

It is generally accepted that if genotoxicity in somatic cells is adequately covered, also the genotoxicity in germ cells is covered. This is based on the assumption that somatic cells are much easier exposed to chemicals in comparison to germ cells. The assumption is confirmed by the fact that there is no unique germ cell mutagen known. From these assumptions and that the classical germ cell genotoxicity tests are rather insensitive, laborious and above all very animal consuming, it is justified to restrict the number of germ cell genotoxicity tests.

The solution introduced in the revised Technical Guidance Document on Risk Assessment (TGD, 2003) seems appropriate. In the TGD, it is stated that expert judgment is needed at this stage to consider whether there is sufficient information to conclude that the substance poses a hazard to germ cells. If this is the case it can be concluded that the substance may cause heritable genetic damage, no further testing is necessary but the chemical is classified as class 2 mutagen. Only if the appraisal of mutagenic potential in germ cells is lacking or inconclusive, additional investigation will be necessary.

If occasionally germ cell mutagenicity testing is essential then the DNA damage should be demonstrated in the progeny. The offspring has to be the target and not the germ cells of the parents. In the case of chromosomal aberrations the dominant lethal test can be performed; for gene mutations the gene mutation test with transgenic animals in which the parents are treated and the mutation frequency is determined in the offspring. Recently, a test determining changes in hyper-variable tandem repetitive regions (minisatellites) that are scattered throughout the chromosomes was suggested to be a proper germ cell mutagenicity test (Somers *et al.*, 2002; Yauk, 2004). Mutations in the tandem repeats in the offspring can be easily detected as fragment lengths polymorphisms after restriction enzyme analysis. As in the gene mutation assay with transgenic mice, mutations will be present in all tissues/cells of the offspring since they originate from the gametes.





## 7. Carcinogenicity testing

Under REACH only occasionally a classical carcinogenicity assay will be performed. Although at all annexes carcinogenicity has to be considered, exclusively at Annex X, a carcinogenicity study may be proposed or even required if the substance has i) a widespread dispersive use or there is evidence of frequent or long term human exposure and ii) the substance is labelled as class 3 mutagen or there is evidence from a repeated dose study that the substance is able to induce hyperplasia and/or pre-neoplastic lesions. REACH also regulates that if a chemical is classified as mutagenic class 1 or 2 a carcinogenicity study will normally not be required. Finally, since cancer is strongly linked with the occurrence of mutations, a classical carcinogenicity may be waived for chemicals which are clearly genotoxic in *in vivo* genotoxicity assays.

It can be assumed that carcinogenicity studies will in future be predominantly performed to demonstrate that a chemical is not carcinogenic, for instance when genotoxicity tests do not result in adequate predictions. In these occasions one has to wonder whether a classical carcinogenicity study is required all the more because of its economic (expensive, duration, animal consuming) and scientific (black box, extrapolation from animal to human, MTP doses) disadvantages (Huff, 1999; Swenberg, 1995; Foran, 1997).

The cell transformation assays (using Syrian Hamster Embryo cells (Berwald and Sach, 1963; Barrett *et al.*, 1984), BALB 3T3 cells (DiPaolo *et al.*, 1972; Kakanuga, 1973) or C3H/10T1/2 cells (Reznikoff *et al.*, 1973) may be a useful screening method, indicating to the carcinogenic potential of substances. However, these tests are *in vitro* tests whereas carcinogenicity is pre-eminently an endpoint that should be studied in laboratory animals.

Transgenic mice, with modifications introduced in genes known to be important in the process of tumour development, could be a promising short term *in vivo* alternative model and potentially provide advantages compared to the lifetime bioassay. Short term carcinogenicity assays using the Tg mouse models, *p53*, Tg.AC, *rasH2*, *Xpa* or *Xpa/p53*, are reviewed in a special issue of Toxicological Pathology (2001), by Pritchard *et al.* (2003) and by de Vries *et al.* (2004).

As mentioned earlier, the waiving of the carcinogenicity assay may cause a problem for risk assessment, done on the results of a carcinogenicity study. Solutions except demanding a carcinogenicity assay are not immediately available. As solutions were mentioned the minimal toxic dose of (sub)-chronic toxicity studies, the TTC concept or the lowest observed effect dose of *in vivo* genotoxicity tests.

Within the REACH strategy (like in present strategies) non-genotoxic carcinogens can be easily missed. Non-genotoxic carcinogens are often recognized in (sub-)chronic studies. However, under REACH these studies are like carcinogenicity studies only very occasionally performed. However, health risk due to non-genotoxic carcinogens may be minimal because the toxicological risk of these compounds may be covered by other toxicological endpoints. Yet, test systems have to be developed to detect these non-genotoxic carcinogens. (Q)SARs or genomics may (in future) be systems with the potential to detect non-genotoxic chemicals.

## 8. Integrated (or Intelligent) Test Systems (ITS)

The idea of integrated (or intelligent) test systems (ITS) is to accelerate evaluations of chemicals and reduce the number of tests. ITS uses non-testing data from other structurally related compounds (read across, category approach) or from *in silico* techniques ((Q)SARs) or combines test systems in an intelligent way. The latter has an additional advantage because it reduces the number of laboratory animals. In short, ITS includes an appropriate evaluation of the existing data including non-testing data (read across, category approach and (Q)SARs), extraction of the information/data gaps and determination of the strategy to close these data gaps.

The usefulness of such techniques varies with the amount and nature of information available, as well as with the specific regulatory questions under consideration. The information may be useful for substances for which no test data are available for instance in assisting setting test priorities. On the other hand, if test data exist, they eventually confirm the results from tests or make a better understanding of test results possible. For substances with a low production volume (< 1 tpa) for which testing is not required this information may be the only option available to establish a hazard profile.

An example for an ITS approach may be the decision to not further testing of a compound which is positive in *in vitro* tests but where toxicokinetic and toxicodynamic studies indicated a very low bioavailability. Further testing would give negative results which are useless since the target tissue was not appropriately exposed.

### 8.1. Prospects of (Quantitative) Structural-Activity Relationships

As (Q)SAR is a most promising and objective non-testing approach it will be discussed further. The (computerized) prediction of the putative genotoxicity and carcinogenicity on the basis of its chemical structure and features and on comparison with known genotoxic chemicals, is based on the assumption that mutagenicity predominantly reflects the ability of a chemical or its metabolite to react with DNA. The idea of structural-activity relationships ((Q)SAR) is not new but has already been promoted by Ashby and Tennant (1991).

In principal (Q)SARs relate biological activities for groups of chemicals to one or more molecular properties (substructures) assuming that they share a common mechanism of action. A distinction is made between Structural-Activity Relationships

(SAR) and Quantitative Structural-Activity Relationships (QSAR). SARs are qualitative relationships between a chemical substructure (structural alert) and the potential of a chemical containing that substructure to exhibit a certain biological effect. (Q)SARs are mathematical models that relate a quantitative measure of chemical structure (*e.g.* a physicochemical property) to a physical property or to a biological effect (*e.g.* a toxicological endpoint like carcinogenesis).

To be able to estimate the value of (Q)SAR data the regulator has to know what the specific (Q)SAR model exactly predicts and what the results mean. Therefore, the (Q)SAR model must have (i) a defined endpoint of regulatory importance, (ii) an unambiguous algorithm giving the regulator insight in the mathematical and statistical procedure and full detail of the training set, (iii) a defined domain of applicability with for the regulator a clear definition of the descriptor and structure spaces and (iv) a mechanistic interpretation. Together with (v) measures of goodness-of-fit, robustness and predictivity, these items form the 'Setubal' principles, developed at a workshop in Setubal (Portugal) which were later adopted by the OECD (CEFIC, 2002; OECD, 2003, 2004). In fact these principles, rephrased into OECD principles, describe the criteria to which a correctly performed (Q)SAR evaluation has to meet. Many (commercial available) (Q)SAR models still have problems to settle to these principles.

In practice, the defined endpoint and insight in the applicability domain is of the highest importance. The applicability domain describes the borders of the (Q)SAR model in which a substance has to fall to get a reliably prediction. The applicability domain is determined by a specific endpoint (*e.g.* Ames test) and entry of data of chemicals which were tested with this test (learning set). For chemicals outside the domain the prediction of the model is of unknown reliability. In other words (Q)SAR predictions are uninterpretable if the applicability domain of the model used is not known. Yet this may be acceptable as long as so much information of the model (applicability domain) is available that the regulator is able to explain the results and knows on the basis of which data the result is positive or negative.

The available models range from methods that identify general 'structural alerts' for genotoxicity such as the Ashby-Tennant super-mutagen molecule (Ashby and Tennant, 1991) to computerized and partly commercially available (Q)SAR models for predicting test results for genotoxic endpoints for closely related structures. The most common computerized models are based on data from the Ames test. Examples of widely used models for this endpoint include an expert system called DEREK, artificial intelligence modules from MULTICASE, the topological system, TOPKAT and the OASIS system which includes a metabolic simulator. Although the conclusions are not consistent, most evaluations indicate that Ames-ruled (Q)SARs predict the Ames test rather well (Ecetoc, 2003; Hulzebos and Posthumus, 2003; Benigni, 2005). There are models to predict many other mutagenicity endpoints. For

example, the Danish EPA have developed with MULTICASE software (Q)SAR systems for *in vitro* endpoints like the chromosomal aberrations (CHO and CHL cells), mouse lymphoma/*tk*, and CHO/*hprt* gene- mutations and for *in vivo* endpoints like mouse micronuclei, rodent dominant lethal mutations and mouse Comets.

When essential features of the information domain are clearly represented, these models may constitute the best predictive tools for estimating a number of mutagenic/genotoxic endpoints. However, quality of reporting varies from model to model and predictivity must be assessed case-by-case on the basis of clear documentation. In the current risk assessment process in the EU for new and existing chemicals, (Q)SARs are used with caution. If well-documented and applicable (Q)SARs are available, they should be used to help to make decisions. In many cases the accuracy of such methods will be sufficient to help, or allow either a testing or a specific regulatory decision to be made. In other cases the uncertainty may be unacceptable due to the severe consequences of a possible error. The reliability of results obtained with (Q)SAR models has to be decided on a case by case basis.

## 8.2. Integration of genotoxicity tests

Genotoxicity test do allow integration into other (toxicity) tests. The most promising is an integration of the *in vivo* micronucleus test using peripheral blood cells, in the 28-day toxicity study. The quantity of blood cells needed for the determination of micronuclei does not hamper the study of haematology in the 28-day study. Moreover, a 28 day treatment period often results in better results in the micronucleus test. In practice there are already some CROs which offer this combination of tests. The performance of a 28-day toxicity study with transgenic (*lacI* or *lacZ*) animals is an extension. Problems may however arise if a part of a tissue is necessary for histopathological evaluation and not enough DNA can be isolated from the remaining tissue. Also other endpoints of genotoxicity can be studied in this integrated test (like unscheduled DNA synthesis).

The germ cell genotoxicity tests which use the progeny as target are pre-eminently appropriate to combine with reproduction toxicology tests. The dominant lethal test strongly resembles the 'One-Generation Reproduction Toxicity Study'(OECD 415) Integration of both the transgenic mouse or the minisatellite test into reproduction toxicity tests has the advantage that there is no time pressure in the F1 generation to collect the tissues for investigation.

For carcinogenicity an intelligent testing strategy also belongs to the possibilities: the classical carcinogenicity test can be combined with a chronic toxicology study. Such combination is very obvious because both tests will only be occasionally performed and a combined OECD guideline already exists (OECD 453).



## 9. Conclusion

The first aim of the present report is to investigate the effect of REACH implementation on genotoxicity and carcinogenicity testing. The REACH requirements to screen chemicals for their genotoxic and carcinogenic properties are adequate as is the concept guidance to translate these requirements into a strategy. Both guarantee a reliable determination of the genotoxic and carcinogenic potency of compounds. In the guidances some flexibility is given to the rather fixed requirements described in the REACH program by indicating to specific circumstances which allow the waiving of experiments. However, adjustments may be needed to comply with a major aim of REACH, a reduction in costs, number of tests and above all the number of laboratory animals. Fortunately, the requirements and the guidances still can be enforced without giving up in essence much on safety assessment. For instance, waiving of further testing for compounds with minimal human exposure that are exclusively weakly positive in the gene mutation test in bacteria or of the carcinogenicity assay for all *in vivo* genotoxic compounds.

Problems with classification and labelling will not occur. Positive results in somatic cells in *in vivo* genotoxicity tests should lead to classification as class 3 mutagen or, if sufficient information is available on germ cell exposure, as a class 2 mutagen without germ cell testing. If *in vivo* testing is not possible classification and labelling should be permitted on *in vitro* data only. Expert judgment together with evidence from (semi-)chronic toxicity, genotoxicity tests or even non-testing data should allow in the absence of a carcinogenicity study the classification of a chemical for carcinogenicity.

REACH may run short on risk assessment for genotoxic carcinogens which relies historically on the results of the carcinogenicity study. After recognized as a genotoxic carcinogen, a dose descriptor and the derived minimal effect level (DMEL) have to be derived. In the absence of a carcinogenicity assay it is unclear how to derive a dose descriptor for risk assessment of genotoxic carcinogens. Solutions may be found in the minimal toxic dose of (sub)-chronic toxicity studies, the TTC concept or the lowest observed effect dose of *in vivo* genotoxicity tests.

The second aim of the present report is to list if at the different levels of testing improvements can be applied.

It appears that it is as yet not possible to determine genotoxicity exclusively on *in vitro* genotoxicity tests due to the high number of false negative and false positive results. It is recommended that a mechanistic justification of all the positive and negative results in genotoxicity tests should be made compulsory. Then many if not all of the false results may be recognized and explained.

The question how often results from *in vivo* tests do change the conclusion based on *in vitro* genotoxicity tests was addressed. Only a first indication of the actual contribution of the *in vivo* tests to detect the genotoxicity of substances was demonstrated. An extensive comparison has to be carried out between the results of *in vitro* and *in vivo* tests covering the same genotoxic endpoint and the same mechanism. Provided that the mechanistic background of the *in vitro* and *in vivo* results is available, a finding that the animal consuming *in vivo* tests only occasionally disagree with the results from *in vitro* tests may justify to evaluate the genotoxic potency of substances exclusively with *in vitro* genotoxicity tests. An appropriate evaluation of the *in vivo* tests itself and of the contribution of these tests in the determination of the genotoxic potency of compounds is anyhow recommendable since under REACH risk assessment of genotoxic carcinogens may in the absence of a carcinogenicity study depend on genotoxicity data, only.

Integrated (or intelligent) test systems (ITS) appear useful tools to accelerate evaluations of chemicals and reduce the number of tests. ITS uses non-testing data from other structurally related compounds (read across, category approach) or from *in silico* techniques ((Q)SARs) or combines test systems in an intelligent way. The information may be useful for substances for which no test data are available for instance in assisting setting test priorities.

Special attention should be given to tests which allow the detection of non-genotoxic carcinogens. Since not only the carcinogenicity assay but also (sub-) chronic studies, predominantly responsible for the detection of non-genotoxic carcinogens, are only very occasionally performed, these substances are easily missed. Fortunately, health risk due to non-genotoxic carcinogens may be minimal because the toxicological risk of these compounds is covered by other toxicological endpoints.

Several recommendations and/or conclusion concerning the strategy were discussed or demonstrated. (1) To determine the intrinsic genotoxic properties of chemicals, a battery of three genotoxicity tests appears preferable but two genotoxicity tests are already sufficient. (2) *In vivo* tests are only then allowed if it is demonstrated that the chemical under investigation reaches the target cells and that the *in vivo* test covers the endpoint that showed positive results *in vitro*. *In vivo* tests should only be performed for substances for which no alternative is available. (3) The absence of tissue restriction and the good performance make the relatively new Comet and transgenic mice assay good alternatives for the classical *in vivo* genotoxicity tests. A positive result in these tests but also in the unscheduled DNA synthesis (indicator) test, is a strong indication that a chemical is a genotoxic carcinogen. (4) Germ cell testing can be reduced to a minimum. It seems justified to consider a substance that is mutagenic to somatic cells and that can reach germ cells as a germ cell mutagen and to classify and label it likewise. (5) If it is possible to derive within a reasonably small range a reliable dose discriminator on other tests than a carcinogenicity test, it may be



justified to waive a carcinogenicity assay for substances classified as class 3 mutagen.  
(6) In these occasions, where under REACH a carcinogenicity study is compulsory, one has to wonder whether a classical carcinogenicity study is required because of its economic and scientific disadvantages.



## 10. Recommendations for further research

The present report provokes at least three recommendations for further research.

- The possibilities for risk assessment of genotoxic carcinogens in the absence of a carcinogenicity assay should be investigated. The question whether it is possible that genotoxicity data instead of or next to data from other (subchronic/chronic) toxicity assay play a role should be addressed.
- A study should be done to explain the false negative and false positive results of (*in vitro*) genotoxicity tests. This will certainly point to the importance of the mechanistic justification of all results of genotoxicity tests. It may lead to make this mechanistic justification compulsory.
- To demonstrate the actual contribution of *in vivo* tests within the strategy to detect the genotoxicity of substances, an extensive comparison has to be carried out between the results of *in vitro* and *in vivo* tests covering the same genotoxic endpoint and the same mechanism. This may ultimately lead to the conclusion that the genotoxic potency of substances can reliably be determined with *in vitro* genotoxicity tests.



## **Acknowledgements**

The author wishes to thank Dr. Ir. B. Hakkert (RIVM, Expertise Centre for Substances), Dr. W.C. Mennes (RIVM, Centre for Substances and Integrated Risk Assessment) and Prof. Dr. H. van Steeg (RIVM, Laboratory for Health Protection Research) for their critical comments regarding this report.



## References

- Adler ID (1984) Cytogenetic tests in mammals. In: Mutagenicity testing: a practical approach. Venitt S, Parry JM (eds), IRL Press, Oxford, Washington DC, pp 275-306.
- Albertini S, van Benthem J, Corvi R, Hoffmann S, Maurici D, Pfuhler S, Vanparys P Report on micronucleus test (MNT) *in vitro*. In preparation.
- Allen JW, Ehling UH, Moore MM, Lewis SE (1995) Germ line specific factors in chemical mutagenesis. *Mutation Res.* 330, 219-231.
- Ames BN, McCann J, Yamasaki E (1975) Methods for detecting carcinogens and mutagens with the Salmonella/mammalian-microsome mutagenicity test. *Mutation Res.* 31, 347-364.
- Ashby J, Lefevre PA, Burlinson B, Penman MG (1985) An assessment of the *in vivo* rat hepatocyte DNA repair assay. *Mutation Res.* 156, 1-18.
- Ashby J, Tennant RW (1991) Definitive relationships among chemical structure, carcinogenicity, and mutagenicity for 301 chemicals tested by the U.S. NTP. *Mutation Res.* 257, 229-306.
- Barrett JC, Hesterberg TW, Thomassen DG (1984) Use of cell transformation systems for carcinogenicity testing and mechanistic studies of carcinogenesis. *Pharmacol. Rev.* 36, 53S-70S.
- Benigni R (2005) Structure-activity relationship studies of chemical mutagens and carcinogens: mechanistic investigations and prediction approaches. *Chem. Rev.* 105, 1767-1800.
- Berwald Y, Sachs L (1963) *In vitro* cell transformation with chemical carcinogens. *Nature* 200, 1182-1184.
- Butterworth BE, Ashby J, Bermudez E, Casciano D, Mirsalis J, Probst G, Williams G. (1987) A protocol and guide for the *in vivo* rat hepatocyte DNA-repair assay. *Mutation Res.* 189, 123-133.
- CEFIC (2002) Report from the workshop on regulatory acceptance of QSARs for human health and environmental endpoints, pp 58. Setubal, Portugal, March 4-6. Report available from [rbu@cefic.be](mailto:rbu@cefic.be).
- GHS, Commission Staff Proposal for a regulation of the European Parliament and of the Council on Classification and Labelling of Substances and Mixtures based on the Globally Harmonised System (2006) Legal Text to the Proposal for a Regulation (corrected 28 August 2006), [http://ec.europa.eu/enterprise/reach/ghs\\_consultation\\_en.htm](http://ec.europa.eu/enterprise/reach/ghs_consultation_en.htm).
- De Vries A, van Steeg H, Opperhuizen A (2004) Transgenic mice as alternatives in carcinogenicity testing: current status. RIVM report 340700001.
- DiPaolo JA, Takano K, Popescu, NC (1972) Quantitation of chemically induced neoplastic transformation of Balb/3T3 cloned cell lines. *Cancer Res.* 32, 2686-2695.

- EC (2001) European Commission, White Paper – Strategy for a future chemicals policy. COM(2001) 88.
- Ecetoc (2003) (Q)SARs: Evaluation of the commercially available software for human health and environmental endpoints with respect to chemical management applications. Technical Report 89.
- EEC (1967) Council Directive 67/548/EEC of 27 June 1967 on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labeling of dangerous substances. O.J. No P196, 16.8.1967, p. 1.
- EU (2003) Proposal for a regulation of the European Parliament and the Council and a proposal for a directive of the European Parliament and the Council. COM(2003) 644.
- EU (2006) Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directive 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC.
- Evans HJ (1976) Cytological methods for detecting chemical mutagens. In: Chemical mutagens, principles and methods for their detection. Volume 4, Hollaender A (ed.), Plenum Press, New York and London, pp 1-29.
- Fenech M, Morley AA (1985) Measurement of micronuclei in lymphocytes. Mutation Res. 147, 29-36.
- Foran JA (1997) Principles for the selection of doses in chronic rodent bioassays. ILSI press, Washington DC.
- GHS (2006) Commission Staff Proposal for a regulation of the European Parliament and of the council on Classification and Labelling of Substances and Mixtures based on the Globally Harmonised System (presented by the commission's services DG Enterprise and Industry and Dg Environment), [http://ec.europa.eu/enterprise/reach/ghs\\_consultation\\_en.htm](http://ec.europa.eu/enterprise/reach/ghs_consultation_en.htm)
- Gossen JA, de Leeuw WJ, Tan CH, Zwarthoff EC, Berends F, Lohman PH, Knook DL, Vijg J (1989) Efficient rescue of integrated shuttle vectors from transgenic mice: a model for studying mutations *in vivo*. Proc. Natl. Acad. Sci. U S A 86, 7971-7975.
- Heddle JA, Salamone MF, Hite M, Kirkhart B, Mavournin K, MacGregor JG, Newell GW (1983) The induction of micronuclei as a measure of genotoxicity. Mutation Res. 123, 61-118.
- Huff J (1999) Long-term chemical carcinogenesis bioassays predict human cancer hazards. Issues, controversies, and uncertainties. Ann. NY Acad. Sci. 895, 56-79.
- Hulzebos EM, Posthumus R (2003) (Q)SARs: Gatekeepers against risk on chemicals? SAR QSAR Environm. Res. 14, 285-316.



- Kakunaga T (1973) Quantitative system for assay of malignant transformation by chemical carcinogens using a clone derived from BALB/c 3T3. *Int. J. Cancer* 12, 463-473.
- Kim BS, Margolin BH (1999) Prediction of rodent carcinogenicity utilizing a battery of *in vitro* and *in vivo* genotoxicity tests. *Environm. Molec. Mutagen.* 34, 297-304.
- Kirkland D, Aardema M, Henderson L, Müller L (2005) Evaluation of the ability of a battery of 3 *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity and relative predictivity. *Mutation Res.* 584, 1-256.
- Kirkland D, Pfuhler S, Tweats D, Aardema M, Corvi R, Darroudi F, Elhajouji A, Glatt H, Hastwell P, Hayashi M, Kasper P, Kirchner S, Lynch A, Marzin D, Maurici D, Meunier J-R, Müller L, Nohynek G, Parry J, Parry E, Thybaud V, Tice R, van Benthem J, Vanparys P, White P (2007) How to reduce false positive results when undertaking *in vitro* genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM Workshop. *Mutation Res.* 628, 31-55.
- Kohler SW, Provost GS, Fieck A, Kretz PL, Bullock WO, Sorge JA, Putman DL, Short JM (1991) Spectra of spontaneous and mutagen-induced mutations in the *lacI* gene in transgenic mice. *Proc. Nat. Acad. Sci. USA.* 88, 7958-7962.
- Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothorst F, Vos JG, Wurtzen G, European branch of the International Life Sciences Institute (2004) Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. *Food Chem. Toxicol.* 42, 65-83.
- Kroes R, Kleiner J, Renwick A (2005) The threshold of toxicological concern in risk assessment. *Toxicol. Sci.* 86, 226-230.
- Lambert IB, Singer TM, Boucher SE, Douglas GR (2005) Detailed review of transgenic rodent mutation assays. *Mutation Res.* 590, 1-280.
- Liber HL, Thilly WG (1982) Mutation assay at the thymidine kinase locus in diploid human lymphoblasts. *Mutation Res.* 94, 467-485.
- Maron DM, Ames BN (1983) Revised methods for the Salmonella mutagenicity test. *Mutation Res.* 113, 173-215.
- Moore MM, DeMarini DM, DeSerres FJ, Tindall KR (1987) Banbury Report 28: Mammalian cell mutagenesis. Cold Spring Harbor Laboratory, New York, New York.
- OECD (2003) Organisation for Economic Co-operation and Development ad-hoc expert group on (Q)SARS: Summary conclusion of the 1<sup>st</sup> meeting and draft work plan. OECD Headquarters, Paris, France.
- OECD (2004) The report from the expert group on (quantitative) structure-activity relationships [(Q)SARs] on the principles for the validation of (Q)SARs. 2<sup>nd</sup> meeting of the ad hoc expert group on QSARs. OECD series on Testing and Assessment, number 49. OECD Headquarters, Paris, France.

- Pedersen F, de Bruijn J, Munn S, van Leeuwen K (2003) Assessment of additional testing needs under REACH. Effects of (Q)SARS, risk based testing and voluntary industry initiatives. JRC Report EUR 20863.
- Preston RJ, Dean BJ, Galloway S, Holden H, McFee AF, Shelby M (1987) Mammalian *in vivo* cytogenetic assays: analysis of chromosome aberrations in bone marrow cells. *Mutation Res.* 189, 157-165.
- Pritchard JB, French JE, Davis BJ, Haseman JK (2003) The role of transgenic mouse models in carcinogen identification. *Environm. Health Perspec.*, 4, 444-454.
- Reznikoff CA, Brankow DW, Heidelberger C (1973) Establishment and characterization of a cloned line of C3H mouse embryo cells sensitive to postconfluence inhibition of division. *Cancer Res.* 33, 3231-3238.
- Sasaki YF, Sekihashi K, Izumiyama F, Nishidate E, Saga A, Ishida K, Tsuda S (2000) The comet assay with multiple mouse organs: comparison of comet assay results and carcinogenicity with 208 chemicals selected from the IARC monographs and U.S. NTP Carcinogenicity Database. *Crit. Rev. Toxicol.* 30, 629-799.
- Somers CM, Yauk C, White PA, Parfett CLJ, Quinn JS (2002) Air pollution induces heritable DNA mutations. *Proc. Natl. Acad. Sci. USA* 99, 15904-15907.
- Swenberg JA (1995) Bioassay design and MTD setting: old methods and new approaches. *Regul. Toxicol. Pharmacol.* 21, 44-51.
- Tice RR, Andrews PW, Singh NP (1990) The single cell gel assay: a sensitive technique for evaluating intercellular differences in DNA damage and repair. *Basic Life Sci.* 53, 291-301.
- Toxicologic Pathology volume 29, supplement. ILSI/HESI Alternatives to Carcinogenicity Testing Project Technical Guidance Document (2003) in support of the Commission Directive 93/67/EEC on risk assessment for new substances, Commission Regulation (EC) no. 1488/94 on risk assessment for existing substances and directive 98/8/EC on the European Parliament and of the Council concerning the placing of biocidal product on the market. EC, Ispra, Italy.
- Van der Jagt K, Munn S, Tørsløv J, de Bruijn J (2004) Alternative approaches can reduce the use of test animals under REACH. Addendum to JRC Report EUR 20863.
- Yauk C (2004) Advances in the application of germline tandem repeat instability for *in situ* monitoring. *Mutation Res.* 566, 169-182.