化学物質の安全性 -*in silico* 評価への挑戦

2011年6月17日(金)

[1](財)食品農医薬品安全性評価センター
[2](独)製品評価技術基盤機構
林 真[1,2], 櫻谷 祐企[2]

1

講演内容

1. 化学物質管理における*in silico* 評価の活用
   - 化学物質審査規制法での活用
   - OECDを中心とした国際的な取組
2. 反復投与毒性を対象とした*in silico* 評価支援システムの開発

2

化学物質審査規制法（化審法）

- 昭和48年10月16日制定され，昭和49年4月施行.
  - ✓世界に先駆けた法律
  - ✓PCBの環境汚染問題による人の健康障害を契機として制定
  - ✓既存化学物質のリスト
    - この時までに日本で生産・輸入されていた化学物質
    - これ以外は新規化学物質として国による審査が必要
- 昭和61年5月　化審法の一部を改正
- 昭和61年12月　スクリーニング毒性試験ガイドライン公表
- 昭和62年3月　その他の毒性試験ガイドライン公表
- 平成15年5月　新化審法が公布
- 平成23年4月 改正化審法施行

3

新規化学物質に係わる主な試験法

分解度試験（微生物による生分解）

濃縮度試験（魚類）

スクリーニング毒性試験　[化学物質が継続的に摂取される場合に人の健康を損なう恐れがあるかどうかについて、下記の試験成績に基づき判定]

変異原性試験

[比較的簡便な短期間の試験により、遺伝毒性、がん原性を予測]

- 細菌を用いる復帰突然変異試験
- ほ乳類培養細胞による変異原試験

28日間の反復投与毒性試験

[動物に被検物質を28日間毎日反復投与したときに現れる毒性を観察する]

生態毒性試験　[化学物質が動植物の生息又は生育に支障を及ぼすおそれがあるものであるかどうかについて、下記の試験成績に基づき判定]

藻類生長阻害試験、ミジンコ急性遊泳阻害試験、魚類急性毒性試験

4

化審法における*in silico* 評価の利用開発

現状は試験優先順位付けや審査時の参考資料としての活用が中心。今後の利用範囲の拡大へ向けて種々の取組がなされている。

試験項目	担当機関	利用モデル等（開発者）
分解度試験	NITE	BIOWIN5,6(US EPA) CATABOL(ブルガス大)
濃縮度試験	NITE	生物濃縮性予測システム(CERI) BCFWIN(US EPA) Baseline Model(ブルガス大) カテゴリーアプローチ(NITE)
変異原生試験	国立衛研	ADME Works(富士通) DEREK(Lhasa) Multicase(Multicase)
生態毒性試験	国環研	KATA(国環研) ECOSAR(US EPA) TIMES(ブルガス大)

5

OECD QSARプログラム*

有害性情報を取得することが必要とされる化学物質の増加（REACH対応等）。

→ 構造活性相関の行政利用を推進することを目的とし、QSARアドホック専門家グループを設立し活動を開始（2003年1月）。

ワークアイテム

① QSARバリデーション原則の確立

② QSAR行政利用のためのガイダンス文書の作成

③ QSARアプリケーションツールボックスの開発

*http://www.oecd.org/document/23/0,3746,en_2649_34365_33957015_1_1_1_1,00.html

6

OECD QSARバリデーション原則*

行政利用の検討を促進するために(Q)SARモデルは以下の情報に関連すべき。

1. エンドポイントの定義
2. 曖昧さのないアルゴリズム
3. 適用範囲の定義
4. 適合度、頑健性、予測性の適切な評価
5. 可能ならば、メカニズムに関する解釈

2004年11月のOECD第37回化学品合同会合にて合意。

*OECD. 2007. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 69, Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models.

7

OECD QSAR Toolbox*

化学物質管理へ携わる人たちへのQSARのavailabilityを向上させることを目的に開発された。カテゴリーアプローチによる評価を支援する機能を主軸とし、各国の実測試験データベースや、記述子算出機能、QSARモデルが集積されている。OECDのHPから無料でダウンロード可能。

Phase 1: Proof of Concept 版
Ver. 1 (2008年3月にリリース)

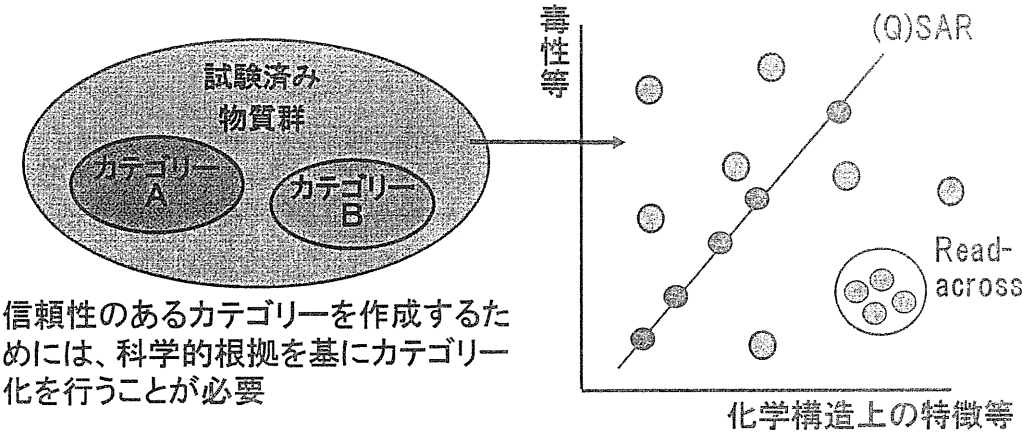
Phase 2: 全ての規制エンドポイントをカバーすることが目標。
Ver. 2 (2010年11月にリリース)
Ver. 3 (2012年10月にリリース予定)
反復投与毒性等、複雑なエンドポイントにも対応させる。

*OECD. 2009. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 102, Guidance document for using the oecd (Q)SAR application toolbox to develop chemical categories according to the oecd guidance on grouping of chemicals.

8

カテゴリーアプローチ*

カテゴリーとは、構造類似性により物理化学的及び毒性学的特性が類似又は規則的なパターンを示すと考えられる化学物質のグループ。構造活性相関[(Q)SARや類推(Read-across)によるデータギャップ補完を行う。



信頼性のあるカテゴリーを作成するためには、科学的根拠を基にカテゴリー化を行うことが必要

透明性の高い評価が可能なことから化学物質管理分野におけるデータギャップ補完手法の主流となっている。

*OECD. 2007. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 80, Guidance on grouping of chemicals.

9

Averse Outcome Pathway (AOP)*

Toxicant	Molecular Initiating Event	Cellular Responses	Organ Responses	Organism Responses	Population Responses
Chemical Properties	Receptor/Ligand Interaction DNA Binding Protein Oxidation	Gene Activation Protein Production Altered Signaling Protein Depletion	Altered Physiology Disrupted Homeostasis Altered Tissue Development or Function	Lethality Impaired Development Impaired Reproduction Cancer	Structure Recruitment Extinction

AOPとは、分子レベルのトリガーとなる反応から、細胞レベル、生レベルのメカニズムを経て、最終的な毒性発現に至るまでの経路を示したもの。現バージョンのツールボックスで対応できていない有害性発現メカニズムが複雑なエンドポイントについては、AOPに基づいてカテゴリー作成するコンセプトがOECDから提案され、最近、精力的に検討がなされている。

*OECD. 2011. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 138, Report of the workshop on using mechanistic information in forming chemical categories.(in press)

10

NEDO/経済産業省委託事業

「構造活性相関手法による有害性評価手法開発」

実施期間：平成19年度～平成23年度

基本計画：化学物質のリスク評価におけるヒト健康影響の評価に際し、安全性試験データがない化学物質に対し、類似化学物質の反復投与毒性試験データやその他の既知見を用いて、カテゴリーアプローチ等の手法により反復投与毒性を推定できるよう必要となる判断材料を評価者（専門家）に提供するデータベース及び評価支援システムを開発する。

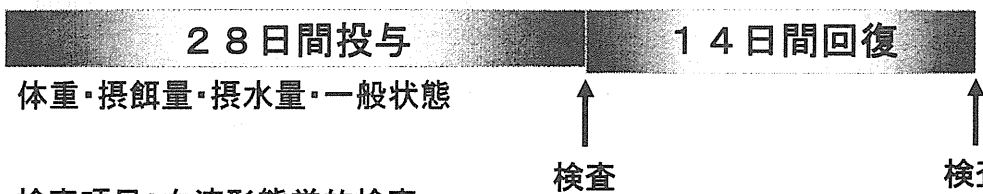
実用的な*In silico* 評価手法が確立されていない反復投与毒性について、*In silico* 評価手法の確立及び評価支援システムを開発することが目的。

11

28日間反復投与毒性試験

目的：動物に被検物質を一定期間毎日反復投与したときに現れる生体の機能及び形態の変化を観察することにより、被検物質の毒性を明らかにすることを目的とする。

齧歯類(原則ラット)

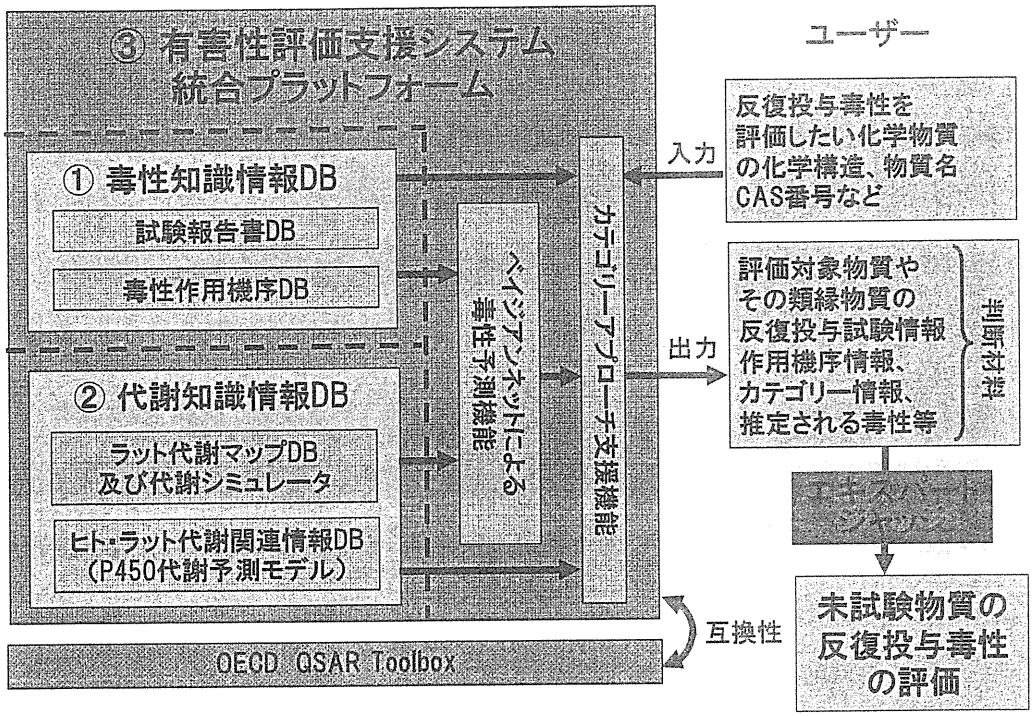


検査項目：血液形態学的検査
血液生化学的検査
神経学的検査
尿検査
病理学的検査　臓器重量
剖検所見
組織学的検査

無影響量(NOEL)
No Observable Effect Level

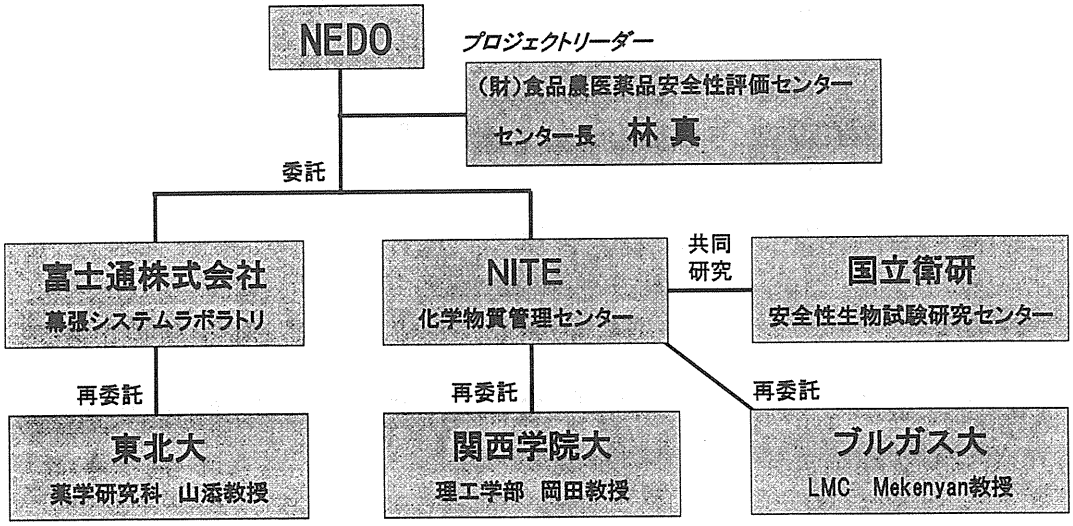
12

開発予定システムの構成



13

実施体制



14

本プロジェクトの特徴

- 国際的に利用されるものを目ざす(OECD (Q)SAR Application Toolboxの開発者をメンバーに加え、OECDと連携しつつ研究開発を実施)。
- 専門家の判断をサポートするためのシステムを構築する(システムが判断を下すのではない)。
- 動態、毒性、病理そして総合的なリスク評価を行う専門家が主導的な役割を果たす(システムの専門家主導ではない)。

15

試験報告書DB (国立衛研、NITE)

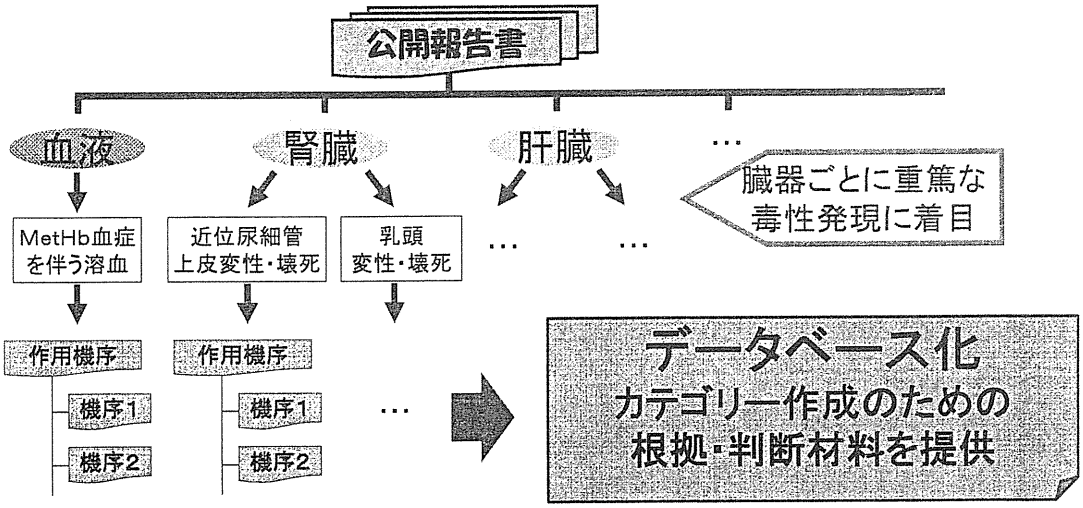
反復投与毒性の毒性学的な類似する物質をカテゴリー化するため、詳細な試験データを含み、かつ、それらを物質間で比較検討できるデータベースを目指した。

- 約500種類の化学物質の反復投与毒性試験の報告書を収集・整理しDB化した。
- 各種検査値や病理組織所見については、群別表をデータベース化 (毒性学専門家が開発した共通フォーマットを使用)。
- 有意差マーク以外に, 専門家により毒性学的と影響が認められた所見(審議会判定等)についてフラグを表示。
- ラボ間での用語のばらつきに対応するため、病理シソーラス(病理学専門家による)を開発し、検索エンジンに反映。

16

毒性作用機序DB（国立衛研、NITE）

反復毒性試験で発現した毒性をベースに物質を選抜し，作用機序を体系的に収集・整理し，データベース化する。



分子、細胞、臓器レベルの作用機序

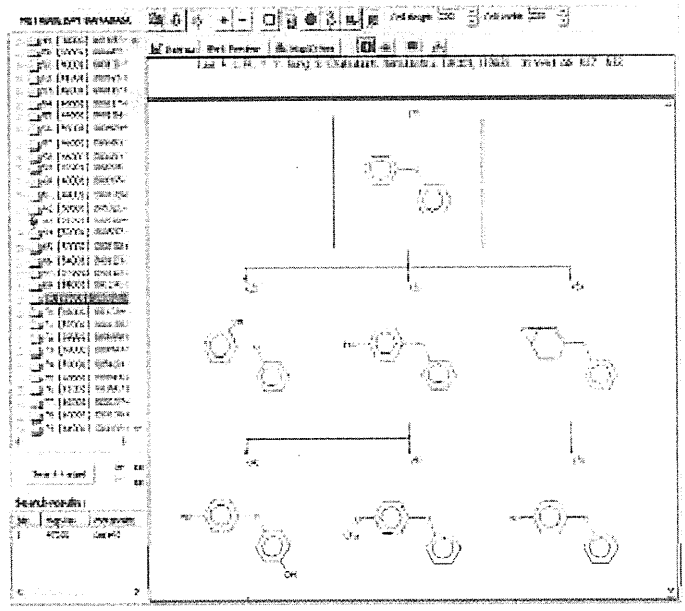
17

ラット代謝マップDB（ブルガス大学、NITE）

反復投与毒性試験報告書がデータベース化された物質（800物質を予定）ついて、既知の代謝経路を収集し、代謝マップを作成し、データベース化。

試験条件や文献情報を併せて登録。

代謝マップの親物質及び代謝物は、部分構造による検索が可能。



18

ヒト・ラット代謝関連情報DB（富士通、東北大学）

反復投与毒性試験報告書がデータベース化された物質について、ヒト/ラット間の毒性発現の種差を検討する際に有用と思われる代謝関連（動態）情報を収集しデータベース化。約100物質について127文献の情報を収載。

代謝関連（動態）情報のデータベース項目

吸収	吸収率 Cmax、Tmax トランスポーターの関与
分布	見かけの分布容積 反復に伴う経時変化 脳→中枢作用 脂肪組織→蓄積 肝臓→酸化抱合代謝 腎臓→尿中排泄 腎臓→タンパク結合 血液よりも高い濃度を示す臓器/器官 トランスポーターの関与

代謝	関連酵素系と分子情報 寄与率 細胞内画分 代謝物 種差 系統差
排泄	排泄率 トランスポーターの関与 種差 系統差
相互作用、酵素阻害、酵素誘導試験の結果	
毒性との関連性	

19

代謝予測手法

実測試験のない物質の代謝情報を補完するため、
収集した代謝情報を、独自の手法により解析し、
化学物質の毒性評価に特化した代謝予測モデルを開発。

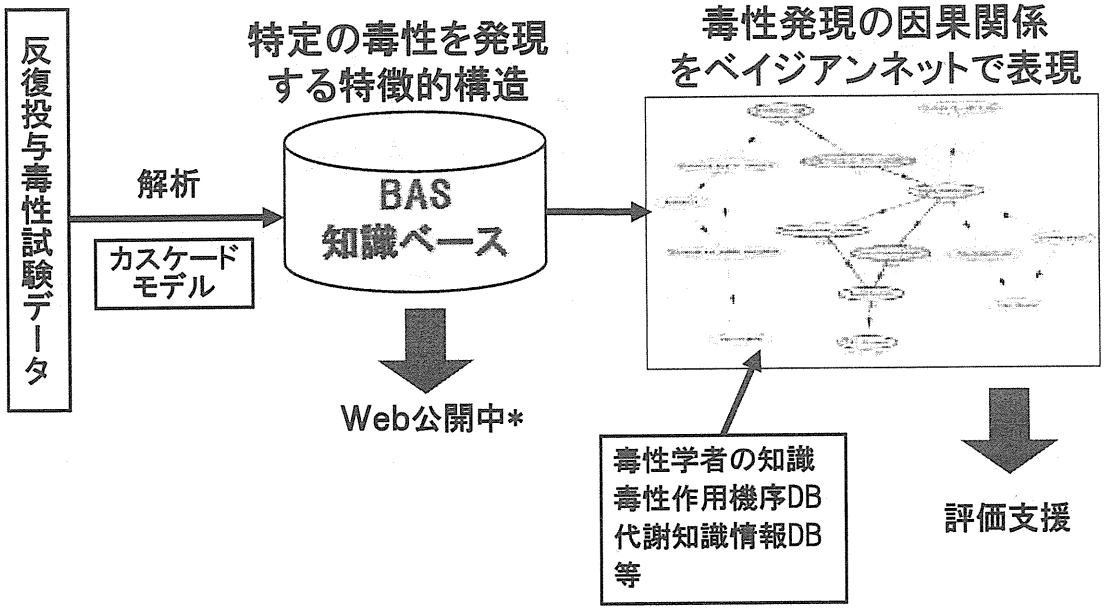
① 代謝シミュレータ（ブルガス大学）
[階層化した反応式と反応確率に基づく予測]
広範囲な物質群に適用可能

② P450代謝予測モデル（東北大学）
[基質の構造に基づく予測]
反応部位を精度良く予測可能
毒性の種差検討への活用が期待される

20

ベイジアンネットによる毒性予測機能

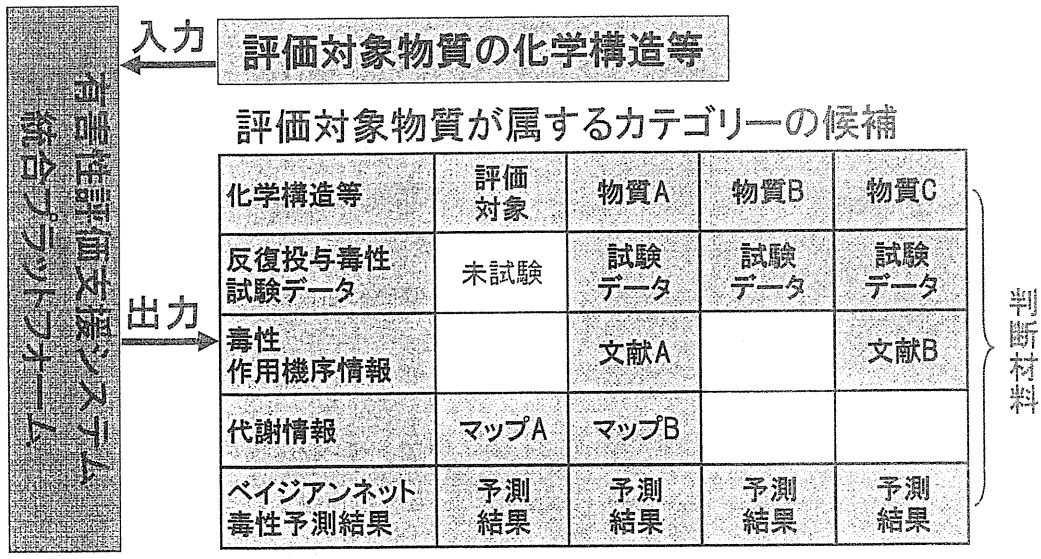
（関西学院大学）



*関西学院大学 BASiC http://www.dm-lab.ws/BASiC/

21

カテゴリーアプローチ支援機能 （NITE、ブルガス大学）



化学構造等	評価対象	物質A	物質B	物質C
反復投与毒性試験データ	未試験	試験データ	試験データ	試験データ
毒性作用機序情報		文献A		文献B
代謝情報	マップA	マップB		
ベイジアンネット毒性予測結果	予測結果	予測結果	予測結果	予測結果

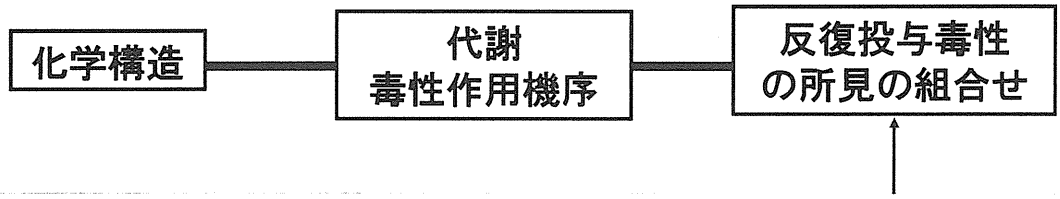
カテゴリーの領域を定義し，システムに登録する。（カテゴリーライブラリー）

22

反復投与毒性のカテゴリー作成の方針

～科学的に根拠のあるカテゴリーを作成するために～

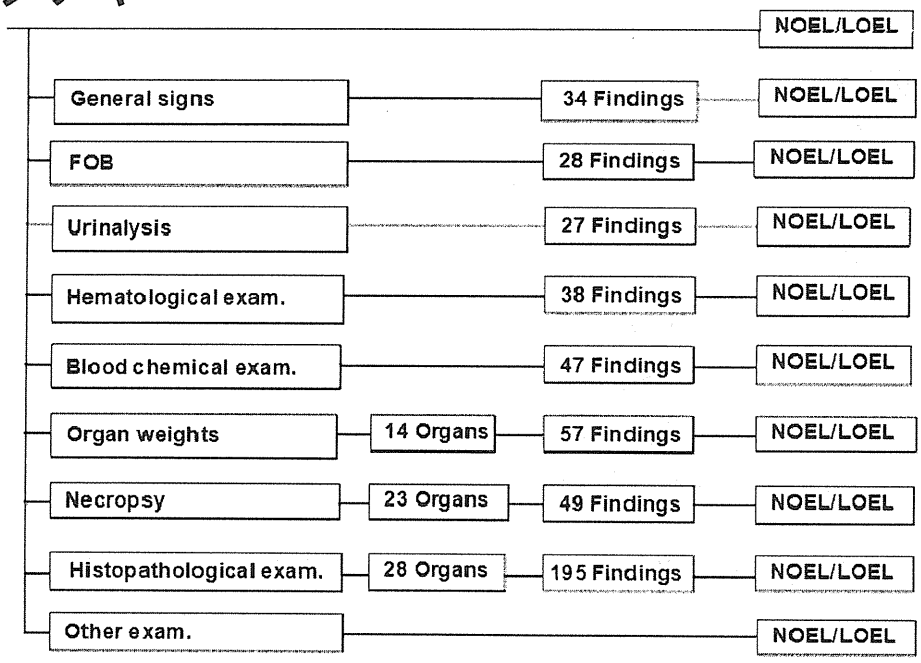
反復投与毒性について，下図のようなAOPを考慮し，同様のAOPと推定される物質群の領域をカテゴリーとして定義した(36種類のカテゴリーを登録)。



反復投与毒性試験の最終的な結果とされている，NO(A)EL，LO(A)ELや標的臓器だけでは，各試験で認められた毒性学的影響の内容を的確に表現できず，物質間で毒性の内容を比較すること(化学構造と毒性を関連付けること)は困難と考え，本研究では，反復投与毒性試験の結果を所見の組合せで表現することとした。

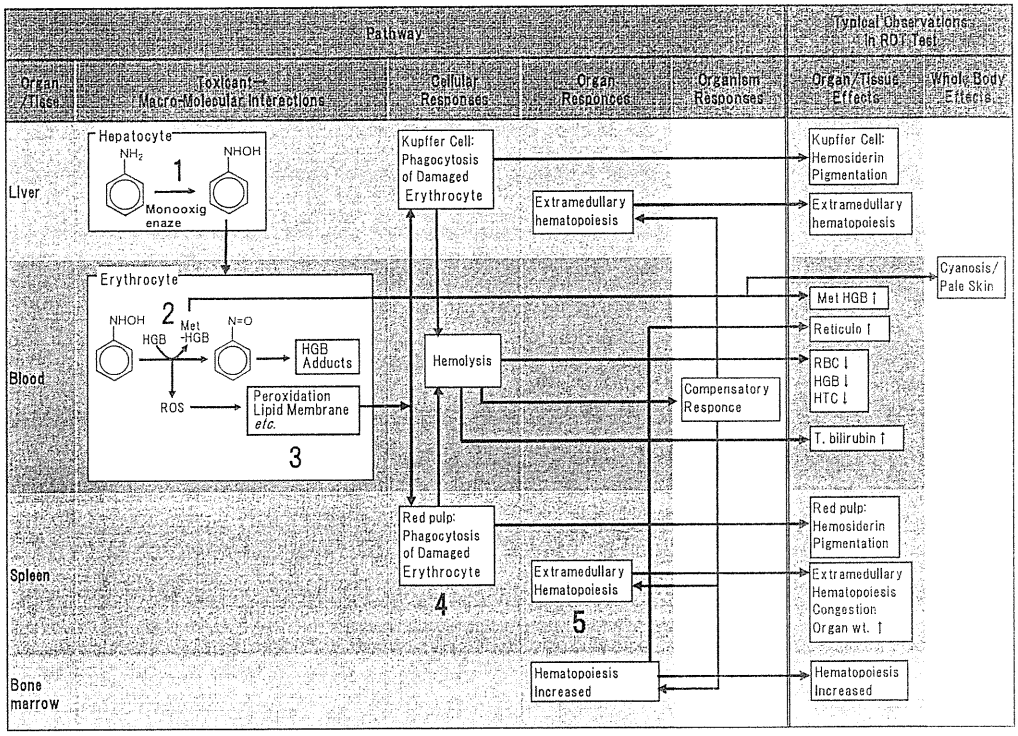
23

反復投与毒性を表現するためのデータストラクチャー



24

反復投与毒性に関するAOPの例



25

まとめ

化学物質の反復投与毒性を、化学構造から推定する際に必要となる情報を収集・整理し、各種データベースと評価支援システムの試作版を完成させた。

毒性・病理学の専門家が開発に深く関与することにより、搭載するデータの質が保証されているだけでなく、病理シソーラスなど、反復投与毒性試験結果を物質間で比較することを支援する機能を備えた、従来にないシステムが開発できた。

収集した情報を活用し、反復投与毒性試験データと化学構造及び毒性作用機序との関係を解析することにより、反復投与毒性についてカテゴリー化の方法論を提案した。

本システムは、平成23年度に外部でトライアルユースを実施する予定。その結果を開発に反映させ、システムを完成させる予定。

26

完成システムの活用について

(1) 化審法での活用を目指す

反復投与毒性に関する初の審査支援ツール

豊富で有用な判断材料を素早く提供することにより的確で効率の良い審査を支援

新規化学物質のデータを随時登録することにより、最新の知見を審査に反映

リスク評価に必要となる判断材料の補完

(2) 事業者の自主的な有害性評価への活用

安全な化学物質を効率良く開発することを支援

（ユーザ独自のデータを追加・解析することも可能）

(3) OECD QSAR Toolboxへの搭載による国際普及

国際的に認められた評価手法とし、加盟国等関係者が同じ手法を用いて評価することができるようにする

27

Mutation Research 723 (2011) 121–128

Contents lists available at ScienceDirect

Mutation Research/Genetic Toxicology and Environmental Mutagenesis

journal homepage: www.elsevier.com/locate/gentox

Community address: www.elsevier.com/locate/mutres

Strategies in case of positive *in vivo* results in genotoxicity testing☆

Véronique Thybaud[a,*], James T. MacGregor[b], Lutz Müller[c], Riccardo Crebelli[d], Kerry Dearfield[e], George Douglas[f], Peter B. Farmer[g], Elmar Gocke[c], Makoto Hayashi[h], David P. Lovell[i], Werner K. Lutz[j], Daniel Marzin[k], Martha Moore[l], Takehiko Nohmi[m], David H. Phillips[n], Jan Van Benthem[o]

[a] *Sanofi-Aventis, Disposition-Safety and Animal Research, Vitry sur Seine, France*
[b] *Toxicology Consulting Services, Arnold, MD, USA*
[c] *F. Hoffmann-La Roche AG, Basel, Switzerland*
[d] *Istituto Superiore di Sanità, Rome, Italy*
[e] *U.S. Department of Agriculture, Food Safety and Inspection Service, Washington, DC USA*
[f] *Health Canada, Ottawa, Ontario, K1A 0K9 Canada*
[g] *University of Leicester, Leicester, UK*
[h] *Biosafety Research Center, Iwata, Japan*
[i] *Division of Basic Medical Sciences, St George's University of London, Cranmer Terrace, London, UK*
[j] *University of Würzburg, Würzburg, Germany*
[k] *Laboratoire de Toxicologie Génétique, Institut Pasteur de Lille, Lille, France*
[l] *National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, Arkansas, USA*
[m] *National Institute of Health Sciences, Tokyo, Japan*
[n] *Institute of Cancer Research, Sutton, UK*
[o] *National Institute for Public Health and the Environment, Laboratory for Health Protection Research, Bilthoven, The Netherlands*

ARTICLE INFO

Article history:
Received 7 September 2010
Accepted 8 September 2010
Available online 17 September 2010

Key words:
Risk assessment
Genotoxicity
In vivo assays
Mode of action
Dose-response

ABSTRACT

At the 2009 International Workshop on Genotoxicity Testing in Basel, an expert group gathered to provide guidance on suitable follow-up tests to describe risk when basic *in vivo* genotoxicity tests have yielded positive results. The working group agreed that non-linear dose-response curves occur *in vivo* with at least some DNA-reactive agents. Quantitative risk assessment in such cases requires the use of (1) adequate data, *i.e.*, the use of all available data for the selection of reliable *in vivo* models to be used for quantitative risk assessment, (2) appropriate mathematical models and statistical analysis for characterizing the dose-response relationships and allowing the use of quantitative and dose-response information in the interpretation of results, (3) mode of action (MOA) information for the evaluation and analysis of risk, and (4) reliable assessments of the internal dose across species for deriving acceptable margins of exposure and risk levels. Hence, the elucidation of MOA and understanding of the mechanism underlying the dose-response curve are important components of risk assessment. The group agreed on the need for (i) the development of *in vivo* assays, especially multi-endpoint, multi-species assays, with emphasis on those applicable to humans, and (ii) consensus about the most appropriate mathematical models and statistical analyses for defining non-linear dose-responses and exposure levels associated with acceptable risk.

1. Introduction

The International Workshops on Genotoxicity Testing (IWGT), in addition to their historical focus on the refinement of genetic toxicology test protocols, have established working groups to recommend appropriate strategies for the use and interpretation of genetic toxicology tests and assessment of the risk of genotoxic exposures. Müller et al. [1] describe the objectives of this IWGT effort, identify areas of focus for the IWGT strategy working groups, and provide initial recommendations for hazard assessment. An IWGT working group has previously provided recommendations on follow-up testing in case of *in vitro* positive results in genotoxicity assays, defined criteria for developing a weight-of-evidence

☆ A report of a working group of the International Workshop on Genotoxicity Testing (IWGT); Basel; Switzerland; August 2009; V. Thybaud; Chair; L. Müller; Co-Chair; J.T. MacGregor; Rapporteur. The opinions in this manuscript are those of the individual authors and do not necessarily represent the policies of their organizations.

* Corresponding author at: Sanofi-Aventis, Vitry-Alfortville Research Center, Disposition-Safety and Animal Research, 13 quai Jules Guesde, 94400 Vitry sur Seine, France. Tel.: +33 01 58 93 35 38.
E-mail address: Veronique.thybaud@sanofi-aventis.com (V. Thybaud).

1383-5718/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.mrgentox.2010.09.002

decision on marginal and non-reproducible results, and has provided a general decision tree for implementing a testing strategy [2]. This general strategy has been extended by a recent working group of the Health and Environmental Sciences Institute, a part of the International Life Sciences Institute (HESI/ILSI), with a focus on interpretation and follow-up testing of positive results *in vitro* [3].

At the August 17–19, 2009, meeting of the IWGT in Basel, Switzerland, a working group was charged with development of recommendations for appropriate follow-up actions when testing results clearly demonstrate genotoxic effects *in vivo*. The objectives of this working group were to develop consensus and to provide recommendations on the following specific topics:

(1) The use of *in vitro* and *in vivo* results in the interpretation and design of *in vivo* assays,
(2) The use of appropriate *in vivo* models for risk assessment,
(3) The evaluation and impact of mode of action (MOA) information, and
(4) Quantitative aspects of the interpretation and use of dose-response information.

This report summarizes the outcome of that working group meeting, including points on which consensus was achieved and those for which further consideration and discussion are needed.

2. Topic 1: Use of *in vitro* and *in vivo* results for risk assessment

In vivo tests are generally considered the most appropriate for quantitative risk estimations. However, they have certain limitations, which are associated both with the current lack of simple assays for relevant endpoints in certain tissues (and the consequent need to interpret surrogate information) and with the difficulty of obtaining mechanistic information in the context of often complex *in vivo* interactions. Thus it was agreed that *in vivo* assays should not be considered in isolation, but that information generated in *in vitro* assays is useful for interpreting the *in vivo* results, defining mode(s) of action, and offering guidance for any additional testing that may be necessary. *In vitro* studies can often be designed to address mechanistic questions that can aid extrapolation to humans. The working group affirmed that when *in vivo* testing is conducted as a follow-up to positive *in vitro* results, the endpoint(s) studied *in vivo* needs to be either the same as that affected *in vitro* or a surrogate demonstrated to be appropriate for predicting the affected endpoint [2] [3]. It was noted that most test systems in genetic toxicology are not specific to a single endpoint and that most DNA-damaging agents affect multiple endpoints, but many chemicals exhibit a predominance of certain types of damage (*e.g.*, point mutations *vs.* chromosomal breaks and rearrangements). Hence, the characteristics of the assays involved (both *in vitro* and *in vivo*) must be taken into account, with recognition of the spectrum of endpoints likely to be affected. Moreover, the selection of tissue(s) *in vivo* should consider pharmaco-/toxicokinetic and pharmaco-/toxicodynamic information about the test material, considering the relevant route(s) of administration. Furthermore, all pertinent toxicological information, including the identification of target organs in sub-acute and sub-chronic studies, should be considered in the design of follow-up *in vivo* genotoxicity studies.

Since absorption, distribution, metabolism, and elimination (ADME) of a compound are integral parts of *in vivo* assays, and ADME extrapolation across species is central to risk assessment, it was acknowledged that the results of the *in vivo* genotoxicity assays should generally have more weight than the *in vitro* assays in genotoxicity risk assessment. Moreover, *in vivo* test selection and design, including selection of tissues for analysis, should be based on appropriate information about anticipated human exposures and account for any known interspecies differences. As *in vivo* testing in genetic toxicology is usually conducted in rodents (mice and rats), well-known limitations exist when extrapolating results from such experiments to the human situation. This is in contrast to other areas of general toxicity testing (*e.g.*, for pharmaceuticals, food additives, and agricultural chemicals), in which extrapolation for humans is normally based on testing in both rodent and non-rodent species. Nonetheless, the working group agreed that if one or more *in vitro* tests are positive, and no measurable genotoxic effects are detected in appropriate *in vivo* endpoints in adequately exposed tissues in relevant animal species, the risk of *in vivo* genotoxic effects can be considered to be negligible. This requires that the follow-up testing *in vivo* is conducted at appropriate doses (*i.e.*, investigated tissues were exposed to sufficient levels of the test material) and that the experimental design is appropriate to show that the *in vitro* effects are not manifested *in vivo*.

The working group also agreed that when genotoxicity is identified in an animal model then an appropriate evaluation of risk in relation to anticipated human exposure should be conducted: *i.e.*, a quantitative risk assessment for the genetic effect should be conducted. The following sections briefly consider the appropriate *in vivo* models that can be used for risk assessment, and then discuss follow-up strategies that can be applied in order to characterize the genotoxic risk for humans.

3. Topic 2: Selection of appropriate *in vivo* models for risk assessment

There was limited discussion of the use of appropriate *in vivo* models for risk assessment. Selection of endpoints and species was not discussed extensively, but there was consensus that transgenic animal models in which neutral reporter genes are used to monitor mutation are acceptable surrogates for naturally occurring genes for assessment of *in vivo* mutagenic risk. The models include transgenic rodent assays with *lacI*, *lacZ* (phage and plasmid), *cII* and *gpt* delta target genes [4]. It appears that such non-transcribed transgenic constructs are useful genetic mutation markers as they are "neutral" and therefore mutations can accumulate during exposure, unlike transcribed genes that may be subject to selective pressure. Thus, the working group agreed that data from transgenic animals with recoverable neutral reporter genes were of comparable quality and predictability for carcinogenicity compared with other standard mutagenicity tests based on endogenous active genes, and that they fill an important need in current regulatory practices (*e.g.*, *in vivo* follow-up testing).

Promising assays include the new *Pig-a* assay [5–12], flow cytometric micronucleus assays [13–27], and *gpt* delta rat and mouse models [28–33]. In particular, the *Pig-a* assay, based on the loss of the glycosylphosphatidylinositol (GPI) membrane anchor of the cell membrane, shows great promise as a high throughput method which, when fully validated, should facilitate the acquisition of data necessary to define *in vivo* dose-response and kinetics relationships that are critical to risk assessment. This assay could easily be coupled with the analysis of micronuclei in peripheral blood. A major advantage of the *Pig-a* and micronucleus assays is that they are conducted using peripheral blood and are therefore relatively non-invasive and can be conducted in any species (including human) [34,35] as part of general toxicity.

The main disadvantage of the *Pig-a* and erythrocyte micronucleus assays is that at present they evaluate only damage in hematopoietic cells, and so are currently not amenable to many target tissues of mutagenesis and carcinogenesis (liver, GI-tract, lung, kidney). Therefore, other assays such as the comet assay or transgenic mutation assays (especially the *gpt* delta model, which

is able to detect both base substitutions and gene deletions) would be needed for the evaluation of *in vivo* genotoxicity in other target organs in which mutations play a significant etiological role in disease. There was consensus that there is a need to continue the development of *in vivo* assays, especially multi-endpoint, multi-species assays, with emphasis on those applicable to samples from human origin. In the spirit of the 3Rs (the Replacement, Reduction and Refinement of the use of experimental animals) in toxicology, integration of *in vivo* genotoxicity assays into 28-day repeat dose toxicity assays or short-term carcinogenicity assays may be an important future direction [4]. Integration into toxicology studies also facilitates comparison of genotoxic responses with other toxicity endpoints and with pharmacokinetic and metabolism information.

4. Topic 3: Quantitative aspects of the interpretation and use of dose-response information

As already reported in the literature [36–39], the working group considered the hypothesis that agents documented to induce genetic damage *via* interaction with non-DNA targets may exhibit a non-linear dose-response relationship with a "threshold" dose below which DNA damage is not expected to occur. For agents that act *via* such non-linear mechanisms, the No Observed Genotoxicity Effect Level (NOGEL) is generally considered an appropriate metric to which additional safety margins may be applied in determining an acceptable safe exposure limit [40]. In such cases, the risk assessment methods applied would be the same as those used for any other toxicological endpoint. For example, in the case of impurities in pharmaceutical agents, the calculation of a permissible daily exposure (PDE) starting with the no observed effect level (NOEL) (or lowest observed effect level (LOEL)) and using five different 'uncertainty factors' has been suggested [40]. The magnitude of the uncertainty factor depends on the degree of certainty for the respective extrapolations from the test systems to the human exposure situation. In the case of pharmaceuticals, an acceptable margin of exposure (MOE) also depends on the benefit of treatment to the exposed patient or population.

The main focus of the working group discussion was the use of quantitative dose-response information to assess the risk of genetic damage due to human exposure to DNA-reactive compounds. In the case of carcinogenicity, the current default assumption is that genotoxic carcinogens that interact with DNA will generally show linear non-threshold dose-responses. However, it has recently been demonstrated that some genotoxic carcinogens that interact with DNA show non-linear dose-response curves with apparent thresholds, *i.e.*, practical thresholds [41–51]. In this paper "threshold" is used to describe a dose below which the incidence of the measured genotoxic effects cannot be distinguished from the background and its associated confidence interval.

Among the non-linear dose-response examples a key case is the recent incident in which the pharmaceutical Viracept was contaminated with ethyl methane sulfonate (EMS), which led to intensive study of the genotoxicity exposure-response relationship for the well-studied genotoxic agent EMS. For this reason, this was selected by the working group as a case study for the quantitative evaluation of the dose-response curves. It was demonstrated in this case that assessment of exposure and response information could be used to define an exposure level, accepted by the European regulatory authority, below which intensive follow-up studies were not considered essential since there is no significant human safety concern. This information, reported at the meeting by Elmar Gocke and Lutz Müller, has now been published ([52,53]; all details in [51]) and is summarized briefly in Appendix A. In their analysis the authors reported that in the case of EMS:

(1) DNA adducts produced by EMS can be repaired error-free,
(2) The existence of several dose groups without any effect below the threshold and their comparison against a large cohort of vehicle controls allows the estimation of a threshold dose and its associated confidence interval,
(3) Assessment of exposure to free EMS in several species appears to be a reasonable basis for human exposure modelling,
(4) It appears that conventional cross-species exposure scaling methods (as used in other areas of toxicology), together with safety margin calculations to balance uncertainties about the exact threshold dose in other species (or other tissues or different age or disease conditions), can be used for risk management for this genotoxic carcinogen.

The direct nature of the genetic damage induced by EMS, which does not involve any major metabolic steps, makes cross-species scaling and risk assessment less complicated than in many other cases, in which metabolic activation or detoxification processes have to be taken into account.

After the Basel meeting, a cancer study using very large numbers of trout exposed to dibenzo[a,l]pyrene was published [54]. The size of this study allowed the determination of statistically significant increases of 1 cancer in 1000 animals. The sensitivity and hence statistical accuracy is more that two orders of magnitude higher than in a usual rodent cancer study with lifetime administration of the test substance.

It was shown that linear extrapolation of the dose-response in the low dose range overestimated the actual cancer risk, and appropriate modelling of the sublinear dose-response curve indicated that the virtual safe dose (VSD; 1 induced cancer in 1,000,000 individuals) was about 1000 fold higher than predicted by linear extrapolation. Notably, the dose-response of the induction of the bulky DNA adducts was close to linear, indicating that at low doses an error-free removal is apparently operative even for the bulky dibenzo[a,l]pyrene adducts. Alternatively, error-free bypass DNA synthesis across the lesion may occur, thereby suppressing the resultant mutations.

Based on the above data and other results recently reported in the literature (*e.g.*, see [41–45] [48] [50] [54–60]) the working group agreed that non-linear response curves and operational thresholds occur *in vivo* with at least some DNA-reactive agents. In other words, some agents will exhibit a "practical threshold": *i.e.*, a dose below which exposure does not add appreciably to existing background rates of DNA damage. Much more data are needed from studies with carefully determined dose-response curves to determine if generalizations across agents are possible. At present, each case requires appropriate data and careful statistical scrutiny. Possible mechanisms/modes of action underlying non-linear dose-response relationships should also be investigated. Consensus is needed on appropriate mathematical models and statistical analyses for characterizing these dose-response relationships and risk levels, and for deriving acceptable margins of exposure. While DNA primary damage can be used for exposure assessment (*i.e.*, as a biomarker of exposure), stable mutations (which are a biomarker of effect) should be given much more weight for risk assessment.

There was consensus that dose and exposure metrics must be justified for each situation of interest, that cross-species extrapolation should consider the same factors that are important for other toxicity endpoints, including relative metabolism, PK differences, surface area scaling, and internal dose, in addition to DNA repair and translesion DNA synthesis differences and relative apoptosis efficiency. Exposure metrics may include the traditional measures of plasma and tissue exposure (C_{max}/AUC).

The working group supported the approach suggested by Lutz and Lutz regarding the analysis of dose-response data for a continuous response variable with background to determine if a threshold

of response is present. This approach, recently published [61], involves definition of the background frequency and variability, followed by a statistical analysis of the data to check whether a fit by a "hockey stick" model is significantly better than a linear regression. If the hockey stick model applies, then the next step consists of a linear regression analysis for the data below the best estimate of the break point in the dose-response curve, estimating the slope of the upper limit of a confidence interval of the linear regression, and calculating the response at the threshold dose. In the context of EMS in Viracept, a 5% error level was proposed [62,63] and a 95% confidence interval was given for the estimate of the calculated threshold dose.

Conventions for unacceptable increases above the existing spontaneous levels need to be established within the scientific community, with consideration of the irreversible nature of mutation induction. The question of a theoretically-expected linear dose-related increase below the threshold dose could be addressed by linear regression of the data below the break point and estimation of an upper limit of the slope. The biological relevance of this slope can then be discussed against the normal variation of background measures in the controls [61]. Other approaches to analysis of thresholds (*e.g.*, [64]) should also be considered, and consensus is still needed about the most appropriate mathematical models and statistical analyses for defining threshold response and exposure levels associated with acceptable risk.

The working group also considered whether genotoxicity data can be used to derive acceptable MOEs, in a manner that is often applied to non-cancer risk factors (*e.g.*, [65]) and sometimes to risk from genotoxic carcinogens [66,67]. To this end, *in vivo* data can be modeled to estimate benchmark dose (BMD_x levels (*i.e.*, dose associated with a defined increase, *x*, of genotoxic damage above background) that could be compared with an estimated human exposure level, as proposed by the European Food Safety Authority (EFSA) for genotoxic carcinogens. As an example an MOE >10,000 relative to the carcinogenic BMD_{10} has been identified as a "low concern" for genotoxic carcinogens by EFSA (2005) [67]. In other words, the threshold of toxicological concern (TTC) levels can be defined for *in vivo* genotoxicants based on benchmark dose level and MOE determinations. Hence, one could determine a permitted daily exposure level with appropriate safety margins for genotoxic carcinogens [53]. For this MOE approach to be applied to *in vivo* genotoxicity data it will be necessary to define the relevant endpoint(s) to be considered and the biologically meaningful increase over background upon which the benchmark dose and safety margins would be chosen. For example, there would be a need for consensus on whether the NOGEL, a particular benchmark dose based on initial response, or other parameter was an appropriate reference exposure parameter for the genotoxic endpoint and also how that exposure metric related to the estimated cancer risk, or other endpoint of concern, for genotoxic agents. Moreover, it can be anticipated that the NOGEL will vary depending on the genetic effect induced and test method applied, and consensus on the selection of relevant endpoints and tests is needed.

The working group felt these approaches should be explored further, but was not able to define the necessary processes at the time of the meeting. The value of such approaches would be that mutagenicity dose-response curves can be determined with far greater precision than carcinogenicity dose-response curves, and so the acceptable margin to avoid genotoxic effects (which might lead to carcinogenicity or other adverse effects) could be determined with much better precision than the acceptable margin to avoid carcinogenicity.

In addition to application of quantitative dose-response information, secondary factors that may modify dose-response relationships were also considered [68–70]. Examples are cell proliferation state, modification or interspecies differences in repair and bypass DNA synthesis capacity or in levels of electrophilic "traps" such as thiols. It was recognized that these factors may be tissue specific and that such factors must be considered when applying quantitative methods to analyse dose-response information.

In summary, there was consensus that quantitative approaches to the assessment of the health risk of exposures to genotoxic agents are necessary when the potential for genotoxic damage that could lead to heritable changes is identified *in vivo*. IWGT will continue to develop recommendations for their implementation.

5. Topic 4: Evaluation and impact of mode of action (MOA) information

Elucidation of MOA of individual compounds is an important component of risk assessment. The better the information about MOA and dose-response relationships, the more certain is the interpretation of dose-response relationships and the determination of an acceptable exposure level in humans. When performing MOA analysis and extrapolating to humans, all available relevant data should be used—not only genotoxicity data.

One example presented was a drug candidate with positive *in vitro* findings that were due to species-specific metabolism that do not occur in humans (Appendix A). Results obtained with the chelating agent nitrilotriacetic acid (NTA) were presented as an example of a rodent nephrocarcinogen with an *in vivo* positive result due to an indirect mechanism of action (Appendix A). When carcinogenicity data are available, genotoxicity should be examined in the target organs for chemical carcinogenesis, using the same species and strains, when possible. Mechanisms underlying the shape of the dose-response curve should be investigated as thoroughly as is feasible both *in vitro* and *in vivo*.

Many chemicals are both mutagens and carcinogens. When conducting an MOA assessment for the induction of the tumors, it is important to consider whether the chemical is actually a mutagenic carcinogen. It should be noted that mutagens should not automatically be assumed to be mutagenic carcinogens. This determination depends on a comprehensive evaluation using a MOA framework and the assessment of key events [71,72].

A strategy for using *in vivo* mutation data to inform cancer MOA was presented. The strategy uses transgenic rodents to evaluate whether a carcinogen can induce mutation in the tumor target tissue. A modified Hill Criteria analysis [73,74] is used to determine whether the induced mutation response is consistent with a mutagenic MOA. This requires an assessment of temporality and dose-response concordance between the mutation dose-response and the tumor dose-response. A case study using riddelliine and dichloroacetic acid (DCA) was presented. Both of these chemicals are mutagens and liver carcinogens. Riddelliine induces mutations in the liver after only a few weeks exposure, while DCA induces mutations in the liver after 60 weeks exposure. A benchmark dose analysis of the mutation and cancer data dose-response curves indicates dose-response concordance for riddelliine but not for DCA. Taken together the temporality analysis and the dose-response concordance analysis for these two chemicals indicate that riddelliine is likely a mutagenic carcinogen, but DCA likely has a different mode of action. The details of this approach are published [75].

It was suggested that future *in vivo* mutation studies to inform MOA should be designed based on the cancer study. Species, dose route, and dose levels should be selected based on the cancer study and should include enough doses, particularly at the lower end of the dose-response curve, to provide an adequate assessment of the dose-response. The design should include chronic exposure and interim sacrifices to provide a dose-response curve at multiple time points. The timing of the interim sacrifices should be based on any

known preneoplastic lesions that occur prior to tumor development. It is possible that the treatment may need to be extended to up to a year, as was the case in the DCA example. Experiments can be designed to evaluate possible MOAs in addition to the induction of mutation.

The extent to which *in vivo* mutagenicity can be associated with adverse effects other than cancer, and the importance of risk assessment of genotoxicity, *per se*, was discussed. In addition to germline mutations that result in well-recognized human diseases, a number of human diseases are caused by *de novo* somatic mutations [76]. More recently, Borlak and co-workers have shown that both somatic and germline mutations result in cardiac septation defects [77–79]. Accordingly, it was affirmed that cancer is not the only adverse health outcome associated with genetic damage, and, therefore *in vivo* genotoxicity should be considered an adverse effect whatever the evidence of carcinogenicity. Data were also presented suggesting that negative carcinogenicity data may not always provide assurance of the lack of genotoxicity *in vivo* in other species or with different exposures. The Maillard reaction product 4-hydroxy-2,5-dimethylfuran-3(2H)-one, negative in a rat carcinogenicity and positive in mutagenicity studies *in vivo* in mouse somatic and germ cells, was discussed as an example [80].

These examples illustrate the need for expert evaluation of all available data to determine the appropriate follow-up investigation that may be necessary for *in vivo* and/or *in vitro* positive genotoxicity data, even when negative carcinogenicity data are available. It was agreed that further review and discussion is warranted before any specific recommendations can be provided on this topic.

6. Conclusions

In conclusion, appropriate models for risk assessment of *in vivo* genotoxicants have been discussed in an IWGT group, and the working group agreed on the following points:

(1) When *in vivo* testing is conducted as a follow-up to positive *in vitro* results, an appropriate experimental design should be used to determine if the *in vitro* effects are manifested *in vivo*, *i.e.*, in adequately exposed tissues in relevant animal species using the same endpoint as that affected *in vitro* or a surrogate demonstrated to be appropriate for predicting the affected endpoint.

(2) Transgenic animal models with recoverable neutral reporter genes are useful for assessing mutagenic activity in different tissues, and are of comparable quality and predictivity for assessment of *in vivo* mutagenic risk as compared to endogenous genes. They therefore fill an important need in current regulatory practice (*e.g.*, *in vivo* follow-up testing). The *Pig*-a assay, flow cytometric micronucleus assays, and *gpt* delta rat and mouse models are promising assays; *Pig-a* and micronucleus assays because they are conducted using peripheral blood and can be conducted in any species, and *gpt* delta model because it is able to detect both point mutations and gene deletions. The comet assay and transgenic mutation assays remain the principal assays allowing the evaluation of *in vivo* genotoxicity in any target organ. There is a need to continue the development of *in vivo* assays, especially multi-endpoint, multi-species assays, with emphasis on those applicable to samples from human origin. Integration of *in vivo* genotoxicity assays into general toxicity assays, such as 28-day repeat dose toxicity assays, is worth considering in the light of its advantages in efficiency, provision of comparative toxicology, pharmacokinetic and metabolic information, and the spirit of the "3Rs" in regulatory toxicology.

(3) Non-linear response curves may occur *in vivo* with non-DNA-reactive and at least some DNA-reactive agents. More data are needed to determine if generalizations across types of agents are possible. Each case requires appropriate data, justified doses and exposure metrics, and careful statistical scrutiny. Consensus is needed on appropriate mathematical models and statistical analyses for characterizing these dose-response relationships and risk levels, and for deriving acceptable margins of exposure. Among the possible approaches are (a) the use of mathematical models and statistical analyses (*e.g.*, "hockey stick" model) to define the background frequency and its variability, and to analyse the dose-response curves, (b) the estimation of the break point in the dose-response curve, *e.g.*, NOGEL with additional safety margins, or BMD to which may be applied an acceptable safe margin of exposure (MOE). In some instances, human exposure levels thought to pose negligible safety concerns can be defined for *in vivo* genotoxicants.

(4) Genotoxic effects *in vivo* generally have more weight than *in vitro* effects in genotoxicity risk assessment, and the absence of measurable *in vivo* effects in target tissues with adequate exposure and metabolic activity indicates that the risk of *in vivo* genotoxic effects can be considered to be negligible in relation to the anticipated human exposure. While DNA primary damage can be used for exposure assessment, stable mutations, *i.e.*, biomarkers of effect, should be given much more weight for risk assessment.

(5) All pertinent toxicological information should be considered in the design of follow-up *in vivo* genotoxicity studies. Elucidation of MOA of individual compounds is an important component of risk assessment, *i.e.*, mechanisms underlying the shape of the dose-response, MOA framework and the assessment of key events, temporality and dose-response concordance between the mutation dose-response, and the tumor dose-response when carcinogenicity data are available.

Conflict of interest

None.

Appendix A. Case examples presented and discussed

Roche Viracept® case

Roche's protease inhibitor nelfinavir mesylate (Viracept®) produced between March 2007 and June 2007 was found to contain elevated levels of EMS, a known mutagen (alkylating agent), leading to a global recall of the drug. EMS levels present in the contaminated drug were predicted not to exceed a dose of ~2.75 mg/day (~0.055 mg/kg/day for a 50 kg patient) based on the daily dose of 2500 mg Viracept/day. As existing toxicology data on EMS did not permit an adequate patient risk assessment, a comprehensive animal toxicology evaluation of EMS was conducted. The general toxicity of EMS was investigated in rats exposed for 28 days. Two studies that assessed DNA damage were performed in mice: chromosomal damage was assessed using a micronucleus assay and gene mutations were detected using the Muta™Mouse transgenic model. In addition, experiments designed to extrapolate animal exposure to humans were undertaken. A general toxicity study showed that the toxicity of EMS occurred only at doses ≥60 mg/kg/day, which is far above the doses received by patients. Studies for chromosomal damage and *lacZ* mutants in mice (in bone marrow and gastrointestinal tract) demonstrated a clear threshold effect with EMS, with no measurable effect at and below 25 mg/kg/day, under 4-week continuous dosing conditions. In the same experiment, a threshold in liver was determined to exist

Table 1
Threshold analysis (hockey-stick model).

Study	Organ	No observed effect level (mg/kg)	Threshold dose (mg/kg)	95% Confidence interval of threshold dose (mg/kg)
Micronucleus test	Bone marrow	80	89.81	56.67–118.25
Muta™mouse	Bone marrow	25	35.45	21.46–45.73
Muta™mouse	Liver	50	51.31	25.67–99.10
Muta™mouse	GI-tract	25	24.51	12.97–38.51

Table 2
Slope analysis for the low dose range.

Study	Organ	Slope at low dose region	95% Confidence interval of slope
Micronucleus test	Bone marrow	−0.10	−0.20 to −0.001
Muta™mouse	Bone marrow	−0.19	−1.19 to 0.81
Muta™mouse	Liver	−0.10	−0.69 to 0.48
Muta™mouse	GI-tract	0.48	−0.96 to 1.92

(The slope is given as number of micronucleated polychromatic erythrocytes out of 4000PCE/mg/kg, and mutant frequency per million cells/mg/kg, for *in vivo* micronucleus test and gene mutation assay in Muta™Mouse, respectively).

at 50 mg/kg/day. A detailed statistical analysis using the approach developed by [61] estimated the 95% statistical confidence intervals for the threshold dose and the slopes below the threshold for the investigated endpoints and organs [62,63]. The confidence in this analysis reflecting a threshold is strengthened by the fact that four dose levels for each organ measured yielded no discernable mutation difference from the control and that three independent control groups were used for the experiment (Tables 1 and 2).

Exposure analysis (C_{max}) in mice, rats and monkeys demonstrated that ~370-fold higher levels of EMS than that ingested by patients are needed to saturate known, highly conserved, error-free, mammalian DNA repair mechanisms for alkylation. Yet, as the half-life of EMS was higher in rats than in mice, and higher in non-human primates than in rats, the calculations of its AUC (area under the exposure-time curve) at the threshold dose of 25 mg/kg/day yielded an AUC-based safety factor of ~28 (*vs.* the C_{max}-based factor of ~370) [81]. Because all mutagenic DNA alkylations seem to be repairable at daily doses up to 25 mg/kg EMS, it can be argued that the C_{max} (which is largely half-life independent) is the main factor for risk assessment in this "EMS in Viracept" case.

In summary, the animal studies suggested that patients who took nelfinavir mesylate (Viracept) with elevated levels of EMS are not at increased risk for carcinogenicity, mutagenicity, or teratogenicity, since mutations are prerequisites for these events. As exposure biomarkers such as adducts on globin or DNA, do generally follow linear dose-response relationships in the case of EMS, these data clearly show that such biomarkers cannot be used for risk assessment or risk management processes in this case but that risk assessment should be based on "fixed" mutational events. Although non-linear behaviour of mutations *in vivo* has been demonstrated previously, these data give the first reliable experimental basis for comprehensive risk management in a low dose exposure scenario.

In vitro effects due to species-specific metabolism

An example of irrelevant *in vivo* positive findings due to an interspecies difference in metabolic capacity was presented. A drug candidate was positive in a comet assay performed on rat stomach and negative in rat liver. This compound is known to be hydroxylated in the stomach and then glucuroconjugated by UDP-glucuronosyltransferase (UDPGT) 1A8. This UDPGT 1A8 isoform is not expressed in the rat gastrointestinal tract [82] while it is highly expressed in human gastric mucosa [83]. *In vitro* assays on isolated mucosa were performed and demonstrated that (1) the glucuro-conjugated metabolite was observed in human gastric mucosa but not in the rat, (2) the hydroxylated metabolite was present in gastric rat mucosa but not in the human, and (3) the hydroxylated metabolite was a direct genotoxic compound in the Ames assay and in the *in vitro* micronucleus assay. It was concluded that the parent compound is positive in the comet assay in rat gastric mucosa but most probably not in human, and that there was no genotoxic concern in human associated with this compound.

Rodent nephrocarcinogen with an indirect mechanism of action

Nitrilotriacetic acid (NTA) induced marked increases in DNA damage after a single oral treatment at high doses in the *in vivo* rodent comet assay on kidney cells at both short (3–6 h) and long-term (22–26 h) sampling times. NTA demonstrated no mutagenic activity in the Ames test but was positive in the *in vitro* micronucleus assay on L5178Y mouse lymphoma cells without and with metabolic activation by aroclor 1254-induced liver or kidney rat S9-mix. An assay on CTLL2/Bcl2 cells coupled to the apoptosis measurement with and without metabolic activation demonstrated a positive response and confirmed the absence of interference of apoptosis. The direct mutagenic activity of NTA was confirmed in the mouse lymphoma *tk*+/− gene mutation assay and in the chromosomal aberrations test on human lymphocytes. However, tested in combination with an excess of Ca^{2+}, NTA gave negative results on L5178Y mouse lymphoma cells, in the *in vitro* comet and in the micronucleus assays, while Ca^{2+} only partly abolished the formation of DNA strand breaks on rat primary kidney cells. The higher sensitivity of renal cells to Ca^{2+} variations could explain the positive response observed *in vivo*. The carcinogenicity of NTA could be a consequence of the intracellular variations of Ca^{2+}, leading to a local and indirect genotoxic mechanism. This suggests that in the case of NTA, a threshold dose may exist beyond which kidney tumor-generating events will be displayed [84].

References

[1] L. Müller, D. Blakey, K.L. Dearfield, S. Galloway, P. Guzzie, M. Hayashi, P. Kasper, D. Kirkland, J.T. MacGregor, J.M. Parry, L. Schechtman, A. Smith, N. Tanaka, D. Tweats, H. Yamasaki, Strategy for genotoxicity testing and stratification of genotoxicity test results—report on initial activities of the IWGT expert group, Mutat. Res. 540 (2003) 177–181.
[2] V. Thybaud, M. Aardema, J. Clements, K. Dearfield, S. Galloway, M. Hayashi, D. Jacobson-Kram, D. Kirkland, J.T. MacGregor, D. Marzin, W. Ohyama, M. Schuler, H. Suzuki, E. Zeiger, Strategy for genotoxicity testing: hazard identification and risk assessment in relation to *in vitro* testing, Mutat. Res. 627 (2007) 41–58.
[3] K.L. Dearfield, V. Thybaud, M.C. Cimino, L. Custer, A. Czich, J. Harvey, S. Hester, J. Kim, D. Kirkland, D. Levy, E. Lorge, M.M. Moore, G. Ouédraogo-Arras, M. Schuler, W. Suter, K. Sweder, K. Tarlo, J. Van Benthem, F. van Goethem, K. Witt, Follow-up actions from positive results of *in vitro* genetic toxicity testing, Environ. Mol. Mutagen. in press.
[4] V. Thybaud, S. Dean, T. Nohmi, J. de Boer, G. Douglas, B. Glickman, N. Gorelick, J. Heddle, R. Heflich, I. Lambert, H. Martus, J. Mirsalis, T. Suzuki, N. Yajima, *In vivo* transgenic mutation assays, Mutat. Res. 540 (2003) 141–151.