

Alternative (non-animal) methods for cosmetics testing: current status and future prospects—2010

Sarah Adler · David Basketter · Stuart Creton · Olavi Pelkonen · Jan van Benthem · Valérie Zuang · Klaus Ejner Andersen · Alexandre Angers-Loustau · Aynur Aptula · Anna Bal-Price · Emilio Benfenati · Ulrike Bernauer · Jos Bessems · Frederic Y. Bois · Alan Boobis · Esther Brandon · Susanne Bremer · Thomas Broschard · Silvia Casati · Sandra Coecke · Raffaella Corvi · Mark Cronin · George Daston · Wolfgang Dekant · Susan Felter · Elise Grignard · Ursula Gundert-Remy · Tuula Heinonen · Ian Kimber · Jos Kleinjans · Hannu Komulainen · Reinhard Kreiling · Joachim Kreysa · Sofia Batista Leite · George Loizou · Gavin Maxwell · Paolo Mazzatorta · Sharon Munn · Stefan Pfuhler · Pascal Phrakonkham · Aldert Piersma · Albrecht Poth · Pilar Prieto · Guillermo Repetto · Vera Rogiers · Greet Schoeters · Michael Schwarz · Rositsa Serafimova · Hanna Tähti · Emanuela Testai · Joost van Delft · Henk van Loveren · Mathieu Vincken · Andrew Worth · José-Manuel Zaldivar

Received: 21 February 2011 / Accepted: 3 March 2011 / Published online: 1 May 2011
© Springer-Verlag 2011

Abstract The 7th amendment to the EU Cosmetics Directive prohibits to put animal-tested cosmetics on the market in Europe after 2013. In that context, the European

Commission invited stakeholder bodies (industry, non-governmental organisations, EU Member States, and the Commission's Scientific Committee on Consumer Safety) to identify scientific experts in five toxicological areas, i.e. toxicokinetics, repeated dose toxicity, carcinogenicity, skin sensitisation, and reproductive toxicity for which the Directive foresees that the 2013 deadline could be further extended in case alternative and validated methods would not be available in time. The selected experts were asked to analyse the status and prospects of alternative methods and to provide a scientifically sound estimate of the time necessary to achieve full replacement of animal testing. In

Sarah Adler, David Basketter, Stuart Creton, Olavi Pelkonen, Jan van Benthem, and Valérie Zuang contributed equally to this work.

Disclaimer: This report represents the opinions of the authors as individual scientists, and should not be taken to represent the positions of any institutions, including those of ECVAM and the other Services of the European Commission.

Electronic supplementary material The online version of this article (doi:10.1007/s00204-011-0693-2) contains supplementary material, which is available to authorized users.

S. Adler
Centre for Documentation and Evaluation of Alternatives to Animal Experiments (ZEBET), Federal Institute for Risk Assessment (BfR), Berlin, Germany

D. Basketter
DABMEB Consultancy Ltd, Sharnbrook, UK

S. Creton
NC3Rs, London, UK

O. Pelkonen
Department of Pharmacology and Toxicology,
University of Oulu, Oulu, Finland

J. van Benthem · A. Piersma
Laboratory for Health Protection Research,
National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

V. Zuang (✉) · A. Angers-Loustau · A. Bal-Price ·
S. Bremer · S. Casati · S. Coecke · R. Corvi · E. Grignard ·
J. Kreysa · S. B. Leite · S. Munn · P. Phrakonkham · P. Prieto ·
R. Serafimova · A. Worth · J.-M. Zaldivar
ECVAM, Institute for Health & Consumer Protection,
European Commission Joint Research Centre, Ispra, VA, Italy
e-mail: valerie.zuang@jrc.ec.europa.eu

K. E. Andersen
Department of Dermatology and Allergy Centre,
Odense University Hospital,
University of Southern Denmark,
Odense, Denmark

A. Aptula · G. Maxwell
SEAC, Unilever, Colworth Science Park,
Sharnbrook, UK

summary, the experts confirmed that it will take at least another 7–9 years for the replacement of the current *in vivo* animal tests used for the safety assessment of cosmetic ingredients for skin sensitisation. However, the experts were also of the opinion that alternative methods may be able to give hazard information, i.e. to differentiate between sensitisers and non-sensitisers, ahead of 2017. This would, however, not provide the complete picture of what is a safe exposure because the relative potency of a sensitiser would not be known. For toxicokinetics, the timeframe was 5–7 years to develop the models still lacking to predict lung absorption and renal/biliary excretion, and even longer to integrate the methods to fully replace the animal toxicokinetic models. For the systemic toxicological endpoints of repeated dose toxicity, carcinogenicity and reproductive toxicity, the time horizon for full replacement could not be estimated.

Keywords Alternative methods · Toxicokinetics · Skin sensitisation · Repeated dose toxicity · Carcinogenicity · Reproductive toxicity

E. Benfenati
Laboratory of Environmental Chemistry and Toxicology,
Istituto di Ricerche Farmacologiche “Mario Negri”,
Milan, Italy

U. Bernauer
Federal Institute for Risk Assessment (BfR), Berlin, Germany

J. Bessems · E. Brandon
Centre for Substances and Integrated Risk Assessment,
National Institute for Public Health and the Environment
(RIVM), Bilthoven, The Netherlands

F. Y. Bois
Chair of Mathematical Modelling for Systems Toxicology,
Compiègne and INERIS, DRC/VIVA/METO, Compiègne
University of Technology, Verneuil en Halatte, France

A. Boobis
Imperial College London, London, UK

T. Broschard
Merck KGaA, Darmstadt, Germany

M. Cronin
School of Pharmacy and Chemistry Liverpool,
John Moores University, Liverpool, UK

G. Daston
Miami Valley Innovation Center, The Procter and Gamble
Company, Cincinnati, OH, USA

W. Dekant
Department of Toxicology, Julius-Maximilians-Universität
Würzburg, Würzburg, Germany

Individual chapters authors contributions:

The authors are solely responsible for the content of the chapter to which they have contributed.

Introduction and overview

Valérie Zuang, Sharon Munn, Joachim Kreysa

1. Toxicokinetics

Olavi Pelkonen (rapporteur), Sandra Coecke (chair),
Sofia Batista Leite, Ulrike Bernauer, Jos Bessems, Esther
Brandon, Frederic Y. Bois, Ursula Gundert-Remy, George
Loizou, Emanuela Testai, José-Manuel Zaldívar

2. Skin sensitisation

David Basketter (rapporteur), Silvia Casati (chair),
Klaus Ejner Andersen, Alexandre Angers-Loustau, Aynur
Aptula, Ian Kimber, Reinhard Kreiling, Henk van Loveren,
Gavin Maxwell, Hanna Tähti

3. Repeated dose toxicity

Stuart Creton (rapporteur), Alan Boobis, Wolfgang
Dekant, Jos Kleinjans, Hannu Komulainen, Paolo

S. Felter · S. Pfulher
Procter and Gamble Company, Central Product Safety,
Cincinnati, OH, USA

U. Gundert-Remy
Department Toxicology, Medical School (Charité),
Berlin, Germany

T. Heinonen · H. Tähti
FICAM, School of Medicine, University of Tampere,
Tampere, Finland

I. Kimber
University of Manchester, Manchester, UK

J. Kleinjans · J. van Delft · H. van Loveren
Department of Health Risk Analyses and Toxicology,
Faculty of Health, Medicine and Life Sciences,
Maastricht University, Maastricht, The Netherlands

H. Komulainen
Department of Environmental Health, National Institute
for Health and Welfare (THL), Kuopio, Finland

R. Kreiling
Clariant Produkte (Deutschland) GmbH, Sulzbach, Germany

P. Mazzatorta
Nestlé Research Center, Vers-chez-les-Blanc,
Lausanne, Switzerland

S. B. Leite
Instituto Tecnologia Química Biológica (ITQB)/Instituto
Biologia Experimental (IBET), Oeiras, Portugal

Mazzatorta, Anna Bal-Price, Vera Rogiers, Greet Schoeters, Mathieu Vinken, Pilar Prieto (chair)

4. Carcinogenicity

Jan van Benthem (rapporteur), Susan Felter, Stefan Pfuhler, Tuula Heinonen, Albrecht Poth, Rositsa Serafimova, Joost van Delft, Emilio Benfenati, Pascal Phrakonham, Andrew Worth, Raffaella Corvi (chair)

5. Reproductive toxicity

Sarah Adler (rapporteur), Thomas Broschard, Susanne Bremer (chair), Mark Cronin, George Daston, Elise Grignard, Aldert Piersma, Guillermo Repetto, Michael Schwarz

Table of contents

Abstract	10
Introduction and Overview	10
Scope	10
Context	11
Evaluation process carried out during 2010	13

G. Loizou
Health and Safety Laboratory,
Mathematical Sciences Unit,
Health Improvement Group,
Harpur Hill, Buxton, UK

A. Poth
Harlan Cytotest Cell Research GmbH,
Roßdorf, Germany

G. Repetto
University Pablo de Olavide, Sevilla, Spain

V. Rogiers · M. Vinken
Department of Toxicology,
Vrije Universiteit Brussel,
Brussels, Belgium

G. Schoeters
Department of Biomedical Sciences,
University of Antwerp, Antwerp, Belgium

M. Schwarz
Institute of Pharmacology und Toxicology,
University of Tuebingen, Tuebingen, Germany

E. Testai
Istituto Superiore di Sanità, Rome, Italy

H. van Loveren
National Institute for Public Health and the Environment
(RIVM), Laboratory for Toxicology,
Pathology and Genetics,
Bilthoven, The Netherlands

Conclusions from each of the five working groups 14

1. Toxicokinetics	14
2. Skin sensitisation	16
3. Repeated Dose Toxicity	18
4. Carcinogenicity	19
5. Reproductive Toxicity	20
Overall Conclusions	21
Future prospects	23
1 Toxicokinetics	26
1.1 Executive Summary	26
1.2 Objectives	28
1.3 Background	29
1.3.1 The TTC concept	29
1.3.2 In vitro toxicokinetics as a key for 1R replacement strategies	31
1.3.3 The relation between kinetics and dynamics for a 1R replacement strategy	32
1.3.4 Importance of analytical methods in the 1R scenario	32
1.3.5 Importance of actual, rather than nominal concentration in the 1R scenario	33
1.4 Strategic considerations of risk assessment of cosmetic ingredients	35
1.5 Available non-animal methods to derive values for absorption, distribution, metabolism and excretion (ADME)	36
1.5.1 Absorption and bioavailability after dermal, inhalatory or oral exposure	36
1.5.1.1 Bioaccessibility models	37
1.5.1.2 Absorption models	37
1.5.1.3 Bioavailability	38
1.5.2 Distribution	39
1.5.2.1 Estimation of plasma protein binding (PPB)	39
1.5.2.2 Estimation of blood-tissue partitioning	39
1.5.2.3 Estimation of substance permeability through specialized barriers	40
1.5.3 Metabolism (Biotransformation)	41
1.5.3.1 <i>In silico</i> approaches	42
1.5.3.2 Metabolic clearance	42
1.5.3.3 Metabolite profile and bioactivation	43
1.5.3.4 Induction assays	44
1.5.3.5 Inhibition assays	44
1.5.4 Excretion	45
1.6 Integrating in vitro and <i>in silico</i> approaches using PBTK modelling	46

- 1.6.1 PBTK models are necessary tools to integrate in vitro and in silico study results 47
- 1.6.2 General description of PBTK models 48
- 1.6.3 Generic applications of PBTK modelling 49
- 1.6.4 Specific applications of PBTK modelling in the case of the 1R for cosmetics 50
- 1.7 Inventory of in vivo Methods Currently Available 52
- 1.8 Inventory of Alternative Methods 53
 - 1.8.1 Currently used in vitro guideline 53
 - 1.8.2 Non-validated human in vitro/in silico approaches 54
 - 1.8.3 Non-validated human in vivo approaches 55
 - 1.8.4 Current developments in model systems 56
 - 1.8.5 Steps or Tests with Novel or Improved Alternative Methods Needed 56
- 1.9 Recommendations 57
 - 1.10 Conclusions 59
- 2 Skin Sensitisation 62
 - 2.1 Executive Summary 62
 - 2.2 Information requirements for the safety assessment of cosmetic ingredients and how this information is currently obtained 64
 - 2.2.1 Introduction/Description of Skin Sensitisation and Mechanisms 64
 - 2.2.2 Inventory of Animal Test Methods Currently Available 65
 - 2.2.2.1 Guinea Pig Maximisation Test (GPMT) 65
 - 2.2.2.2 Buehler Guinea Pig Test 65
 - 2.2.2.3 Mouse Local Lymph Node Assay (LLNA) 66
 - 2.2.3 Information Supplied by these Tests and its Use for Risk Assessment 66
 - 2.2.4 Current Risk Assessment 67
 - 2.2.5 Non-Animal Tools to Inform the Cosmetic Industry for Risk Assessment 69
 - 2.2.5.1 Bioavailability 70
 - 2.2.5.2 Mechanistic Chemistry and *In Chemico* Reactivity (mechanistic step 2) 71
 - 2.2.5.3 Epidermal Inflammatory Responses (*mechanistic step 3*) 75
 - 2.2.5.4 Human Skin Equivalents (Reconstituted Tissue Models) (*mechanistic step 3*) 76
 - 2.2.5.5 Dendritic Cell Responses (*mechanistic step 4*) 77
 - 2.2.5.6 Keratinocyte/DC Co-culture Systems (*mechanistic step 4*) 79
 - 2.2.5.7 Dendritic Cell Migration (*mechanistic step 5*) 80
 - 2.2.5.8 T Cell Responses (mechanistic step 6) 80
- 2.3 Identified areas with no alternative methods available and related scientific/technical difficulties 81
- 2.4 Summary of alternative methods currently available and foreseeable time to achieve full replacement of the animal test(s) 82
- 2.5 Conclusions/Summary 82
- 3 Repeated dose toxicity 83
 - 3.1 Executive Summary 83
 - 3.2 Introduction 85
 - 3.3 Current repeated dose toxicity methodology for the safety assessment of cosmetic ingredients 86
 - 3.4 Current availability and status of alternative methods for repeated dose toxicity 88
 - 3.4.1 Introduction 88
 - 3.4.2 (Q)SARs and in silico modelling 90
 - 3.4.3 In vitro models 93
 - 3.4.3.1 Hepatotoxicity 94
 - 3.4.3.2 Nephrotoxicity 96
 - 3.4.3.3 Cardiovascular toxicity 97
 - 3.4.3.4 Neurotoxicity 98
 - 3.4.3.5 Immunotoxicity and myelotoxicity 101
 - 3.4.4 Omics and imaging technologies 103
 - 3.4.5 Strategies to reduce, refine or replace the use of animals 104
 - 3.5 Challenges for the development of alternative approaches for quantitative risk assessment of cosmetic ingredients 110
 - 3.5.1 Quantitative Risk Assessment 110
 - 3.5.2 Limitations of in vivo studies related to quantitative risk assessment 111
 - 3.5.3 Limitations of in vitro studies with specific cell types 113
 - 3.5.4 Importance of understanding of mode of action and toxicity pathways in the development of alternative approaches 115
 - 3.5.5 Current initiatives to develop alternative approaches for repeated dose toxicity 117
- 3.6 Conclusions 118
- 4 Carcinogenicity 120
 - 4.1 Executive summary 120
 - 4.2 General considerations 121
 - 4.2.1 Introduction 121
 - 4.2.2 Information Requirements for the carcinogenic Safety Assessment of Cosmetics Ingredients until March 2009 (Ref. SCCP Notes of guidance) 123

- 4.2.3 Implications for carcinogenic safety assessment after the 7th amendment 124
- 4.2.4 Assessment of genotoxic carcinogens 125
- 4.2.5 Assessment of non-genotoxic carcinogens 127
- 4.3 Inventory of Alternative Methods Currently Available 129
 - 4.3.1 Non-testing methods 129
 - 4.3.1.1 Quantitative structure–activity relationship (QSAR) 130
 - 4.3.1.2 Read-across and grouping of chemicals 133
 - 4.3.1.3 Threshold of Toxicologic Concern (TTC) approach 136
 - 4.3.2 In vitro methods 139
 - 4.3.2.1 Classical genotoxicity tests 139
 - 4.3.2.2 In vitro Micronucleus test in 3D human reconstructed skin models (RSMN) 141
 - 4.3.2.3 In vitro Comet assay in 3D human reconstructed skin models 142
 - 4.3.2.4 GreenScreen HC assay 143
 - 4.3.2.5 Hens egg test for micronucleus induction (HET-MN) 144
 - 4.3.2.6 Cell transformation assay 145
 - 4.3.2.7 In vitro toxicogenomics 147
 - 4.3.3 In vivo methods (Reduction/refinement) 148
 - 4.3.3.1 In vivo genotoxicity tests 149
 - 4.3.3.2 Transgenic mouse models 150
 - 4.3.3.3 In vivo toxicogenomics 151
- 4.4 Identified Areas with no Alternative Methods Available and Related Scientific/Technical Difficulties 152
- 4.5 Conclusions 154
- 5 Reproductive Toxicity 156
 - 5.1 Executive Summary 156
 - 5.2 Introduction 157
 - 5.2.1 Complexity of the Reproductive Cycle 157
 - 5.2.2 Alternatives for Reproductive Toxicity Testing 157
 - 5.3 Information Requirements for the Safety Assessment of Cosmetic 158
 - 5.4 Inventory of Animal Test Methods Currently Used for the Evaluation of Developmental and Reproductive Toxicity 159
 - 5.4.1 OECD Test Guideline 414: Prenatal Development Toxicity Study for the Testing of Chemicals (OECD, 2001a) 160
 - 5.4.2 OECD Test Guideline 415: One-Generation Reproduction Toxicity Study (OECD, 1983) 161
 - 5.4.3 OECD Test Guideline 416: Two-Generation Reproduction Toxicity (OECD, 2001b) 161
 - 5.4.4 OECD Test Guideline 421: Reproduction/Developmental Toxicity Screening Test (OECD, 1995) 162
 - 5.4.5 OECD Test Guideline 422: Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test (OECD, 1996) 162
 - 5.4.6 OECD Test Guideline 426: Developmental Neurotoxicity Study (OECD, 2007c) 163
 - 5.4.7 OECD Test Guideline 440: Uterotrophic Bioassay in Rodents: A short-term screening test for oestrogenic properties (OECD, 2007d) 163
 - 5.4.8 OECD Test Guideline 455: The Stably Transfected Human Estrogen Receptor- α Transcriptional Activation Assay for Detection of Estrogenic Agonist-Activity of Chemicals (OECD, 2009c) 164
 - 5.4.9 Draft OECD Test Guideline Extended One-Generation Reproductive Toxicity Study (OECD, 2009d) 164
 - 5.5 Inventory of Alternative Methods 164
 - 5.5.1 Developmental Toxicity 165
 - 5.5.1.1 Whole Embryo Tests 165
 - 5.5.1.2 The Micromass Test 168
 - 5.5.1.3 Pluripotent Stem Cell-based in vitro Tests 168
 - 5.5.2 Placental Toxicity and Transport 170
 - 5.5.2.1 The Placental Perfusion Assay 170
 - 5.5.2.2 Trophoblast Cell Assay 170
 - 5.5.3 Preimplantation Toxicity 171
 - 5.5.3.1 Male Fertility 171
 - 5.5.3.2 Female Fertility 173
 - 5.5.4 In vitro Tests for Assessing Effects on the Endocrine System 174
 - 5.5.4.1 Ishikawa Cell Test 174
 - 5.5.4.2 Cell Proliferation Based Assays for Testing Estrogen Activity 175
 - 5.5.4.3 Receptor Binding Assays 175
 - 5.5.4.4 Transcriptional Tests 176
 - 5.5.4.5 Tests Assessing Steroidogenesis 177
 - 5.5.5 Application of In Silico Techniques to Reproductive Toxicology 178

- 5.5.5.1 Existing Data 178
- 5.5.5.2 Grouping/Category Formation 179
- 5.5.5.3 Structure–Activity Relationships (SARs) 181
- 5.5.5.4 QSARs 181
- 5.5.5.5 *In Silico* Approaches for Endocrine Mechanisms of Action 183
- 5.5.5.6 Current Status of *In Silico* Approaches for predicting Reproductive Toxicity 183
- 5.6 Identified Areas with no Alternative Methods Available and Related Scientific/Technical Difficulties 185
 - 5.6.1 Approaches for alternative Testing 185
 - 5.6.2 General limitations of *in vitro* methods for reproductive toxicity testing 185
 - 5.6.3 The Testing Strategy as the Future Driving Force 186
 - 5.6.4 Retrospective Analyses to Select Critical Endpoints 186
 - 5.6.5 Towards the Definition of Novel Testing Paradigms 187
 - 5.6.6 Time Schedule for Phasing out *in vivo* Reproductive Toxicity Testing 187
- 6. Disclaimer 188
- 7. References 188

Introduction and overview

Scope

This report provides the findings of a panel of scientific experts tasked with assessing the availability of alternative methods to animal testing in view of the full marketing ban foreseen in 2013 for cosmetic products and ingredients tested on animals in Europe.

Context

There has been a continuous effort at EU level to find alternative approaches which avoid testing on animals wherever possible. Whenever replacement is not possible, the development of methods which use fewer animals or cause least harm to the animals is supported. This ‘Three Rs Principle’ (replacement, reduction and refinement of animal use) is present in all relevant EU legislations.

In early 2003, the 7th amendment to the European Union’s Cosmetics Directive (76/768/EEC) was adopted. It stipulated an immediate end to animal testing in the EU for cosmetic products and a complete ban of animal testing for ingredients by 11 March 2009, irrespective of the

availability of alternative methods. The animal testing ban was reinforced by a marketing ban on all cosmetic ingredients or products tested for the purposes of the Directive outside the EU after the same date. The only exception related to animal testing for the more complex toxicological endpoints such as repeated dose toxicity, reproductive toxicity and toxicokinetics, for which the deadline was set to 11 March 2013, respecting that alternatives for these human health (-related) effects would not be available by 2009. The Directive foresees that the 2013 deadline could be further extended in case alternative and validated methods would not be available in time. Further to the adoption of the 7th Amendment, the European Commission was tasked with reporting regularly on progress and compliance with the deadlines as well as possible technical difficulties in complying with this ban.

Already in 2003, after consultation with the main stakeholders in the field, the Commission established a panel of 75 scientific experts, drawn from various stakeholder bodies.¹ The panel was requested to establish timetables² for phasing out animal testing for an agreed number² of human health effects of concern. The European Centre for the Validation of Alternative Methods (ECVAM), hosted by the Institute for Health and Consumer Protection of the European Commission’s Joint Research Centre, was asked to coordinate this activity.

On the basis of an inventory of the available alternative methods in the respective toxicological areas, the experts estimated the time required to bring the methods to regulatory acceptance. The results of the panel’s work were published in a scientific journal in 2005.³ The most favourable outlook in terms of the time estimated for full animal replacement (5 years or less) was in the area of skin irritation. Test methods for skin corrosion, skin absorption/penetration and phototoxicity were already adopted into legislation at that time. Prospects for the mid- to long-term (over 5 and up to 15 years) included the areas of eye irritation, acute toxicity, skin sensitisation, genotoxicity and mutagenicity, and toxicokinetics and metabolism.

Areas for which a replacement could, according to the experts, not even be estimated on the basis of state-of-the-

¹ Representing industry, animal welfare and consumer associations, governmental laboratories, and academia.

² 11 in total: acute toxicity; skin irritation and corrosion; eye irritation; skin sensitisation; skin absorption and penetration; subacute and subchronic toxicity; genotoxicity and mutagenicity; UV-induced toxic effects; toxicokinetics and metabolism; carcinogenicity; and reproductive and developmental toxicity.

³ Alternatives to Laboratory Animals (ATLA), volume 33, supplement 1, July 2005.

art techniques available in 2005 included: photosensitisation, sub-acute and sub-chronic toxicity, carcinogenicity, and reproductive and developmental toxicity.

Between 2003 and 2010, considerable efforts were made to accelerate the availability of suitable and appropriate alternative tests. For example, the European Commission funded research programmes in the area of alternatives in the region of 150 million € over the FP6 and FP7 framework programmes.

Many other international research programmes and industry initiatives have contributed to the efforts to find alternative methods over the past years. Furthermore, ECVAM, existing since 1991, has invested considerable time and resources in coordinating and promoting the development, validation and use of alternative methods.

Since 2005, ECVAM has provided annual technical reports on the progress made on the development, validation and regulatory acceptance of alternative methods as an input to the Commission's yearly report on progress and compliance with the deadlines of the Directive.⁴ These reports confirmed that the estimates made by the experts in 2005 were broadly accurate, in that full replacement has been achieved in the areas of skin irritation and corrosion, skin absorption and penetration, and phototoxicity. There are also regulatory accepted methods for the identification of severe eye irritants and good progress is being made towards a testing strategy for the full replacement of the Draize (rabbit) eye irritation test. Nevertheless, full replacements do not yet exist for eye irritation, genotoxicity and acute toxicity, whereas the testing and marketing ban is already in force for these endpoints.

In 2011, the Commission is called upon to review the situation regarding the technical difficulties in complying with the 2013 ban and inform the European Parliament and the Council and propose adequate measures, if necessary. In this context, in 2010 the Commission decided to conduct a similar exercise as in 2005 on the current status of development of alternative methods and future prospects for evaluating repeated dose toxicity (including skin sensitisation and carcinogenicity), toxicokinetics and reproductive toxicity. ECVAM was asked again to coordinate the work.

Evaluation process carried out during 2010

In 2010, the European Commission invited stakeholder bodies (including industry, non-governmental organisations, EU Member States, and the Commission's Scientific Committee on Consumer Safety—SCCS) to nominate scientific experts for each of the five toxicological areas of

⁴ Latest available report: "ECVAM technical report on the Status of Alternative Methods for Cosmetics Testing (2008–2009)", 2010.

concern, i.e. toxicokinetics, repeated dose toxicity, carcinogenicity, skin sensitisation and reproductive toxicity. From these suggested experts, thirty-nine were selected with a view to have a balanced coverage of expertise needed. The selected experts (see Annex 1) were invited to participate in one of the five working groups according to their expertise, acting in their personal capacity and not representing any organisation or interest group. Each working group, chaired by ECVAM staff, was required to analyse the specific types of information provided by the animal test methods used for safety assessment and to compare this with the information that could be derived from appropriate alternative methods. Concerning the latter, the experts were asked to provide realistic estimates of the time required for the development of such methods (where they did not already exist) to a level approaching readiness for validation; i.e. meeting the criteria to enter pre-validation.⁵

The time needed for (pre-) validation and regulatory acceptance of alternative methods was not to be included because, on the basis of estimates already produced by ECVAM in 2005⁶ and taking account of recent progress in this field, it can be estimated that validation would require 2–3 years and regulatory acceptance an additional 2–5 years. Therefore, another 4–8 years need to be realistically added to the time estimates for the research and development efforts before regulatory risk assessment would become feasible without any animal experiments. It is also important to note that these estimates were made assuming that optimal conditions are met. This means that all necessary resources (technical, human, financial and coordination) are available at all times in the process and that the studies undertaken have successful outcomes.

The draft report of each working group was published on the Commission's Europa website⁷, and comments from the general public were invited from 23 July 2010 to 15 October 2010. During that time, some thousand factual and editorial comments were received which were carefully considered by the working groups and integrated into the final reports, when and where appropriate.

Conclusions from each of the five working groups

The resulting full reports of the five working groups are presented in the subsequent chapters while main findings are given here below. The time estimates provided by the

⁵ Pre-validation is a small-scale inter-laboratory study to assess the readiness of a test for inclusion in a formal large-scale validation study; Curren et al. (1995).

⁶ Eskes and Zuang (2005, 228 pp).

⁷ Under the http://ec.europa.eu/consumers/sectors/cosmetics/files/pdf/animal_testing/consultation_animal_testing_en.pdf.

experts are based on 2010 as a starting point, as the evaluation was carried out in the course of 2010.

Toxicokinetics

Toxicokinetics is the endpoint that informs about the penetration into and fate within the body (i.e. its toxicokinetics) of a toxic substance, including the possible emergence of metabolites and their absorption, distribution, metabolism and excretion (ADME). While toxicokinetics is an intrinsic part of an *in vivo* animal study, when developing an alternative approach based on *in vitro* studies, toxicokinetics becomes an absolutely crucial and indispensable first step in translating the observations *in vitro* to the human *in vivo* situation.

Toxicokinetics can also inform on the need for further testing based on bioavailability considerations. Hence, for safety assessment of a cosmetic ingredient, testing for systemic toxicity is only necessary if the ingredient penetrates into the body following dermal, oral, or inhalation exposure and if internal exposure potentially exceeds critical levels, i.e. the internal Threshold of Toxicological Concern (TTC, see Annex 2). The TTC concept aims to establish a human exposure threshold value below which there is a very low probability of an appreciable risk to human health, applicable to chemicals for which toxicological data are not available and based on chemical structure and toxicity data of structurally related chemicals.

Determining the internal TTC, a novel concept explained in Annex 2, is of highest importance. Unfortunately, the currently available data are too sparse to allow derivation of any internal TTC.

Knowledge of toxicokinetics is also needed to estimate the possible range of target doses at the cell or tissue level that can be expected from realistic external human exposure scenarios to cosmetics. This information is crucial for determining the dose range that should be used for *in vitro* testing.

Kinetics in the *in vitro* system and dose–response information is also crucial to translate *in vitro* results to the (human) *in vivo* situation. They have key importance for a further development of alternative testing for systemic toxicity capable of full replacement of animal experiments. Therefore, full integration of kinetic expertise into design and execution of toxicity testing and risk assessment is essential.

For the proper design and performance of *in vitro* studies aiming at determining systemic toxicity effects, it is important to include kinetic and analytical aspects in the *in vitro* test protocols (see the framework proposed in Annex 2). Analytical aspects are further important to measure biomarkers as indicators for toxic effects in *in vitro* tests.

Toxicokinetic modelling is currently seen as the most adequate approach to simulate the fate of compounds in the human body. However, high-quality data are needed as input for these models. These data should and can be generated with non-animal studies with *in vitro* or *in silico* approaches that allow quantification of specific dose–response curves.

In conclusion, for most kinetic data, non-animal methods are indeed available or at an advanced development stage. However, alternative methods are lacking for predicting renal and biliary excretion, as well as absorption in the lung, and the experts estimated that at least another 5–7 years (2015–2017) would be needed to develop appropriate models. The full replacement of current animal toxicokinetics tests, linking the results from *in vitro*/*in silico* methods with toxicokinetic modelling, will take even more time, and no specific timeline for this could be given by the experts, but would be clearly beyond 2013.

Skin sensitisation

Skin sensitisation is the toxicological endpoint associated with chemicals that have the intrinsic ability to cause skin allergy, termed allergic contact dermatitis (ACD) in humans. The mechanisms at the basis of the induction of skin sensitisation are rather complex but relatively well understood and involve the following key steps: skin bioavailability, haptentation (binding to proteins), epidermal inflammation, dendritic cell activation, dendritic cell migration and T-cell proliferation. Skin sensitisation (and ACD) appears only after repeated exposure.

Predictive testing to identify and characterise substances causing skin sensitisation historically has been based on animal tests. While all tests are able to differentiate between non-sensitisers and sensitisers, one test, the local lymph node assay (LLNA), is regarded as more capable to predict the relative potency of skin sensitising chemicals, i.e. the chemical's relative power/strength to induce skin sensitisation. This is crucial since skin sensitisers are known to vary by up to 5 orders of magnitude with respect to their relative skin sensitising potency, and only potency information enables the establishment of safe levels for human exposure to chemicals that cannot be regarded as not having any skin sensitising potential.

In recent years, non-animal alternative methods have been developed and evaluated to identify skin sensitisation hazard potential. Although some publications indicate that hazard identification might indeed be possible using these methods, at the moment, none of these tests has been formally validated.

However, due to the complexity of the endpoint, it has been anticipated that no single non-animal approach could generate the potency information that would be required to

fully assess the safety of sensitisers and to allow prediction of a safe level of human exposure to chemicals that carry a known level of skin sensitising hazard. Consequently, a range of non-animal test methods that address the above-mentioned key mechanisms involved in skin sensitisation would be necessary to yield an alternative measure of skin sensitiser potency. However, at present, it is not possible to predict which combinations of non-animal information will be needed before risk assessment decisions could be exclusively based on non-animal testing data with sufficient confidence for the vast majority of cosmetic product exposure scenarios.

It is recognised, however, that data from non-animal test methods developed to identify skin sensitisation hazard potential could be applied to risk assessment decision-making. For example, if the absence of a skin sensitisation hazard was reliably identified, no additional information on potency would need to be generated.

In conclusion, on the basis of the above, the experts agreed that by 2013, no full replacement of animal methods will be available for skin sensitising potency assessment. The most positive view of timing for this is another 7–9 years (2017–2019), but alternative methods able to discriminate between sensitisers and non-sensitisers might become available earlier. However, hazard identifying non-animal tests in isolation will not be sufficient to fully replace the need for animal testing for this endpoint. In estimating the full replacement timeline, the experts of this working group, unlike the experts of the other working groups, included the time required, typically 2–3 years, for demonstration that a non-animal test method is robust, relevant and delivers useful information. Furthermore, the indicated timeline is based upon the premise that predictive non-animal test methods would be available for each mechanistic step. On the other hand, it is unlikely that information on every mechanistic step will be required to inform all risk assessment decisions. Therefore, it is expected that the scientific ability to inform skin sensitisation decisions without animal test data for some ingredients and exposure scenarios should be feasible ahead of 2017–2019.

Repeated dose toxicity

The term repeated dose toxicity comprises the general toxicological effects occurring as a result of repeated daily dosing with, or exposure to, a substance for a part of the expected lifespan (sub-chronic exposure) or, in case of chronic exposure, for the major part of the lifespan of the experimental animal. The onset and progression of this toxicity is influenced by the interplay between different cell types, tissues and organs, including the concomitant contribution of toxicokinetics, hormonal effects, autonomic

nervous system, immunosystem and other complex systems. Current repeated dose toxicity studies provide information on a wide range of endpoints because changes in many organs and tissues are taken into account. They allow evaluation of an integrated response and its quantitative (dose–response) aspects, making its replacement very challenging.

To date, alternative methods have been developed mainly with the aim of producing stand-alone methods predicting effects in specific target organs. However, for the purpose of quantitative risk assessment, an integrated approach, e.g. based on the understanding of the mode of action and perturbation of biological pathways leading to toxicity, is needed. Integrative research efforts considering interactions between different biological tissues and systems, which would be more representative of the situation in the human body, have only recently been initiated.

In addition to *in vitro* techniques, initial attempts of computer-based modelling techniques have suggested the feasibility of developing models providing meaningful predictions of chronic or repeated dose toxicity. At present, there are only a few such models available. Also in the recent years, “omics” technologies have been applied to *in vitro* models for the purpose of understanding, and ultimately predicting toxicity and these technologies hold considerable promise.

The experts concluded that these and other methods under development may be useful for identifying the potential adverse effects of substances (hazard identification) to a limited number of target organs or for obtaining mechanistic information, but none of them is currently seen as appropriate for providing all information needed for quantitative safety assessment related to repeated dose toxicity of cosmetic ingredients.

For quantitative risk assessment, better and more scientific knowledge on exposure, toxicokinetics, dose response and mechanisms of toxicity are needed. Approaches must be developed for combining and interpreting data on multiple targets obtained from a variety of alternative methods and on the extrapolation between exposure routes. Although efforts are on the way to use models for prospective quantitative risk assessment for repeated dose toxicity, additional efforts are necessary to develop improved biokinetic models which would be able to correctly estimate the impact of the distribution over time and level of the repeated external exposure and resulting internal dose. Such models are also needed for extrapolating from *in vitro* to *in vivo* and for understanding of dose responses, so that the *in vitro* data can be applied for quantitative risk assessment.

In conclusion, in view of this non-exhaustive list of scientific challenges, full replacement of the animal tests currently used for repeated dose toxicity testing will not be

available by 2013. No estimate of the time needed to achieve full replacement could be made by the experts, because this will also depend on the progress of (basic) research and development, adequate prioritisation, funding and coordination of efforts, including those to translate basic research results into practical and robust alternative test methods that allow adequate safety assessment and risk management.

Carcinogenicity

Carcinogenesis is a long-term, highly complex process that is characterised by a sequence of stages, complex biological interactions and many different modes of action. It is recognised that even for one chemical substance, the mode of action can be different in different target organs, and/or in different species. Such complex adverse effects are to date neither fully understood nor can they be completely mimicked by the use of non-animal tests.

The 2-year cancer assay in rodents is widely regarded as the “gold standard” to evaluate the cancer hazard and potency of a chemical substance. However, this test is rarely done for cosmetic ingredients. Rather, a combination of shorter-term *in vitro* and *in vivo* studies have been used including *in vitro* and *in vivo* genotoxicity assays to assess genotoxic potential and repeated dose (typically 90-day) toxicity studies to assess the risk of non-genotoxic chemicals. It is clear that the animal testing ban under the 7th amendment of the Cosmetics Directive will have a strong impact on the ability to evaluate and conduct a quantitative risk assessment for potential carcinogenicity of new cosmetic ingredients. This impact is not only due to the ban on the cancer assay itself, but mainly to the ban on *in vivo* genotoxicity testing, any repeated dose toxicity testing, and other tests such as *in vivo* toxicokinetics studies and *in vivo* mechanistic assays.

Although several *in vitro* short-term tests which are at different stages of development and scientific and regulatory acceptance are available beyond the standard *in vitro* genotoxicity assays to support conclusions on cancer hazard identification, the available *in vitro* short-term tests are not sufficient to fully replace the animal tests needed to confirm the safety of cosmetic ingredients. Those that are available are focused on hazard evaluation only and cannot currently be used to support a full safety assessment with adequate dose–response information. However, for some chemical classes, the available non-animal methods might be sufficient to rule out carcinogenic potential in a weight of evidence approach.

In conclusion, taking into consideration the present state of the art of the non-animal methods, the experts were not in a position to suggest a timeline for full replacement of animal tests currently needed to fully evaluate carcinogenic

risks of chemicals. Although a timeline for full replacement cannot be developed, clearly the timeline is expected to extend beyond 2013.

Reproductive toxicity

Reproductive toxicity is probably the most difficult endpoint to be replaced, since it has not only to provide an understanding of the many mechanisms and their interactions which are essential for male and female fertility but also an understanding of the development of the entire human being during its prenatal life. It is therefore not yet possible to estimate the impact that disturbing single or multiple of these mechanisms could have on the entire reproductive process including the normal postnatal development. Only animal models are hitherto accepted as adequately representing the complexity and providing an assessment of the complex interaction of chemicals on the reproductive system. This complexity explains the slow progress in developing and implementing alternatives for reproductive toxicity safety assessments.

However, in the last decades, some non-animal *in vitro* tests have been developed and validated that address specific aspects, such as embryotoxic- or endocrine disrupting effects, of the overall reproductive system. So far, however, only selected mechanisms, which lead to reproductive toxicity, can be mimicked *in vitro*. The available tests are used as screening tools or for providing additional or supporting mechanistic information, but no single alternative method or set of methods is yet available that could replace the current animal tests used for assessing reproductive toxicity of chemical substances. On the other hand, regulators take account of data generated with these tests as additional and supporting evidence when carrying out safety assessments.

A promising way forward is the use of recently established comprehensive databases in which toxicological information derived from standardised animal experiments is collected. These databases will allow the identification of the most sensitive targets/target mechanisms of reproductive toxicants. A mapping exercise is then needed to identify for which endpoints promising and reliable alternative testing methods are already available and which missing “building blocks” for complete integrated testing strategies still need to be developed. Use of advanced *in vitro* technologies, including stem cells, should be considered for generating these missing methods.

In conclusion, given the complexity of the system that needs to be modelled, and the many endpoints and their interactions which must be addressed, the experts estimated that this will need more than 10 years to be completed and thus will clearly not be achieved by 2013.

Overall conclusions

From the extensive reports of the working groups and the discussions held with the experts involved, a number of general conclusions can be drawn:

Despite the considerable efforts and progress which have been made since the last status report produced by ECVAM in 2005, the scientific basis to fully replace animal testing for toxicokinetics and the systemic toxicological endpoints is still not fully established and will need additional time beyond 2013.

All reports of the working groups, as summarised above and shown in detail in the following chapters, highlighted that, at present, animal tests are still necessary for carrying out a full safety assessment of cosmetic ingredients with regard to the critical toxicological endpoints (i.e. toxicokinetics, repeated dose toxicity incl. skin sensitisation and carcinogenicity, and reproductive toxicity). They underlined the critical lack of alternative methods which would be able to provide, for these endpoints, a similar basis for safety assessment as the current animal tests, i.e. the lack of tests which are not only able to differentiate between substances that have a certain adverse effect or not (hazard), but also allow an estimation of the strength of this effect (i.e. potency, dose/response) in order to establish a safe exposure below which no adverse effect would be expected.

The working group experts confirmed that it could take at least another 7–9 years for the replacement of the current *in vivo* animal tests used for the safety assessment of cosmetic ingredients for skin sensitisation. This confirms the forecast of up to 15 years made in the 2005 status report. However, the experts were also of the opinion that alternative methods may be able to give hazard information, i.e. to differentiate between sensitisers and non-sensitisers, ahead of 2017. This would, however, not provide the complete picture of what is a safe exposure because the relative potency of a sensitiser would not be known.

For toxicokinetics, the time frame was 5–7 years to develop the models still lacking to predict lung absorption and renal/biliary excretion, and even longer to integrate the methods to fully replace the animal toxicokinetic models. For the systemic toxicological endpoints of repeated dose toxicity, carcinogenicity and reproductive toxicity, the time horizon for full replacement still could not be estimated at this time.

However, this does not mean that no progress has been made in the last 5 years. On the contrary, the extensive descriptions of both *in vitro* and *in silico* models in the tables and accompanying texts of the five endpoint-specific chapters of this report indicate that many methods are in development.

However, even if these methods were validated in the coming years, it is not yet clear in most cases how to combine the non-animal information in a way that would provide the confidence required for basing risk assessment decisions exclusively on non-animal (testing) data.

The greater rate of progress in finding alternatives in the area of skin sensitisation is because this is a systemic toxicological effect for which the mechanisms are relatively well understood, which increases the possibility to develop non-animal methods to model or measure certain elements of that mechanism. Accordingly, a number of mechanism-specific tests for skin sensitisation are being developed, some of which are already in the validation process by ECVAM. Although the current *in vitro* tests have been mainly designed to differentiate between skin sensitisers and non-sensitisers, some may contribute to the determination of potency which is crucial for risk assessment. How the information from those tests will be exploited to predict potency is currently under investigation, and additional time will be needed before such information can be used with sufficient confidence to fully replace the need for animal testing for risk assessment.

For general systemic toxicity, including reproductive toxicity and carcinogenicity, the many mechanisms leading to toxicity are only poorly defined at present. Single cell-based assays are not able to represent adequately the complex interplay between different cell types in a specific organ, and the involvement of mediators released by other tissues such as the immune, the inflammatory or the endocrine system. It is also unclear how representative cells *in vitro* are of the behaviour of cells *in vivo*. More work is needed to understand this issue. Having said this, many *in vitro* and *in silico* methods have been developed or are being developed to assess the impact of chemicals on specific elements of the complex mechanisms and biological systems which are at the basis of systemic toxicity.

Future prospects

As mentioned several times previously, toxicokinetics was identified as an indispensable element for future non-animal testing approaches, because it is needed to determine the internal dose that reaches, at a given external exposure, the target cells or organs. This information is very important for the translation of *in vitro* studies to the human *in vivo* situation. This can be achieved through the use of physiologically based toxicokinetic (PBTK) modelling to estimate from measured *in vitro* concentrations the relevant exposure at organism level. A framework for carrying out this *in vitro* to *in vivo* translation is outlined in a simple manner in Annex 2 and further described in the chapter on toxicokinetics.

From the risk assessment perspective, the Threshold of Toxicological Concern (TTC) concept, using exposure as a driver for the need for testing or not, was considered as a potentially useful approach by several working groups (e.g. working groups on toxicokinetics, repeated dose toxicity and carcinogenicity), who also provide a detailed discussion of it in their respective chapters. It could, for example, form part of integrated testing strategies, as a pragmatic assessment tool to avoid (or reduce) the need for *in vivo* or *in vitro* testing. Work to further develop such risk assessment approaches, for example tiered approaches which incorporate information on skin and lung penetration, to assess potential internal exposure, is also recommended. This work should include consideration of how data on consumer exposure can be used to determine the kind of information that is required and hence the need for testing. If exposure levels could be determined below which no adverse effect is expected, regardless of the substance's potential hazard, animal testing could be avoided for all uses where the expected exposure remain below the TTC. Such a tiered approach is outlined in a proposal for a framework for risk assessment without animal testing described in Annex 2 to this chapter. While this concept is already embraced for food additives and contaminants and also pesticides, it is not yet widely applied for cosmetics. The experts suggested that it is worth consideration. However, appropriate databases would need to be developed.

Recent advances in cell-based research including use of stem cells, and the development of two-dimensional and three-dimensional cell (co)-cultures, are facilitating the development of much more sophisticated structures more similar to tissues in the body. In addition, the advent of “omics”-based technologies (i.e. genomics, metabonomics, proteomics), which can measure the impact of a chemical on gene and protein expression and metabolism within the cell, should be able to indicate potential pathways by which a chemical may act upon the body. Computer-based *in silico* methods are also being applied to model the interactions of different biological systems and to link the toxic effects of a chemical to its structure and other descriptors. These new technologies hold a lot of promise for the future development of a more predictive risk assessment, based on a better understanding of how toxic substances reach the target cells and organs and perturb critical biological pathways. The integration of some of these methods in so-called intelligent or integrated testing strategies is seen as the most promising way forward. While each of the methods alone may not be able to generate all required information, their combination might provide a sufficient basis for a complete safety assessment.

It should be borne in mind that the animal tests which need to be replaced are not always relevant to predict

human risk and have inherent limitations as well. However, in terms of complexity, the animals are still considered to be the closest approximation to the human body that is currently available. The challenge to reflect the systemic effects with *in vitro* and/or *in silico* models is high, and it is unrealistic to think that it will be feasible to model the entire mechanisms of action for a particular toxicological endpoint within the short- to mid-term time horizons. It is therefore essential to identify those mechanisms of action which drive the toxicity and to focus the development of alternative test methods on these. Exploring developments made in biomedical research and basic biological sciences will be of central importance for making swift progress. With this in mind, it is not unrealistic to expect that in the future, human health safety assessments can be realised without recourse to animals as model organisms, and that our understanding of the nature and level of risk will even improve. However, this will require time, as well as continuing research, development and innovation in this field. In any case, the momentum for developing alternative methods and testing strategies should be maintained. Research and development activities in the field of non-animal testing, both in the public sector (EC framework programmes and national research programmes) and in the private industry sector, have already yielded many promising methods and approaches, and these activities should be further stimulated and encouraged.

For repeated dose toxicity, the recently initiated research activity that is cofunded by the Commission and the Cosmetics Industry under the European Commission's FP7 HEALTH Programme is expected to contribute significantly, since it is designed to contribute to the development of alternative methods that can form building blocks in the integrated approach needed for quantitative risk assessment. This project, known as SEURAT “Safety Evaluation Ultimately Replacing Animal Testing”, started on 1 January 2011 for 5 years and is composed of six complementary research projects. These six projects will closely cooperate in a cluster with a common goal and join over 70 European universities, public research institutes and companies.

In the short term, providing a “toolbox” of well-defined test methods with established reliability and relevance for particular purposes, e.g. clearly addressing well-defined mechanisms of toxicity, could support the development of integrated testing strategies with an ultimate aim of completely replacing animal testing, even for systemic toxicity endpoints.

It is noteworthy that the progress made in the development of non-animal approaches will also be useful in other regulatory contexts besides that of cosmetics, e.g. those related to safety of food, and chemicals in consumer products and the environment. Hence, this will support the

overall aim to replace, reduce and refine animal experiments. While full replacement is not yet accomplished or possible by 2013, the working groups nevertheless agreed that there is a potential for partial replacement strategies, to reduce the number of animals used in the shorter term and underlined that opportunities for reduction and refinement should be pursued within existing approaches wherever possible.

Toxicokinetics

Executive summary

1. Given the scenario of the development of cosmetic products based on non-animal testing strategies non-in vivo animal testing strategies (1R)⁸, toxicokinetics becomes the essential and central body of information.
2. Information on toxicokinetics under 1R is indispensable to address three major issues:
 - (A) *Development and design of more efficient testing strategies*: As a key starting point for any toxicological testing, it is essential to know whether a compound and/or its metabolites will be bioavailable by one of the relevant uptake routes. Only in cases where a cosmetic ingredient is bioavailable following dermal, oral or inhalation exposure, further tests on systemic and not just local toxicity would be necessary.
 - (B) *In vitro–in vivo extrapolation*: To relate toxicodynamic information from non-in vivo animal testing (1R) to real-life situation relevant for humans, i.e. to transform an in vitro concentration–effect relationship into an in vivo dose–effect relationship. In this respect, the role of in vitro biokinetics is crucial to translate a nominal in vitro concentration to the actual level of cell exposure producing the observed effects. For the proper design and performance of in vitro studies, it is important to include kinetic and analytical aspects in the in vitro test protocols.
 - (C) *Identification of clearance rates and the role of metabolites*: For the in vitro dynamics experiments, it is essential to know whether the cell or tissue under human exposure conditions is exposed to the parent compound and/or its metabolites. This information is required upfront and can be obtained from toxicokinetic

alternative methods that identify the main metabolites and the clearance rates of the parent compound and/or its metabolites.

3. Under 1R, toxicokinetic studies can make use of the updated OECD 417 (July 2010) which also comprises in vitro (e.g. use of microsomal fractions or cell lines to address metabolism) and in silico (toxicokinetic modelling for the prediction of systemic exposure and internal tissue dose) methods (OECD 2010a).
4. Physiologically based toxicokinetic (PBTK) models are ideally suited for the integration of data produced from in vitro/in silico methods into a biologically meaningful framework and for the extrapolation to in vivo conditions.
5. Sensitive, specific and validated analytical methods for a new substance and its potential metabolites will be an indispensable step in gathering data for quantitative risk assessment.
6. A whole array of in vitro/in silico methods at various levels of development is available for most of the steps and mechanisms which govern the toxicokinetics of cosmetic substances. One exception is excretion, for which until now no in vitro/in silico methods are available; thus, there is an urgent need for further developments in this area. Also there is a lack of experience for absorption through the lung alveoli, which would also make this a priority item for research and development given the fact that this route of exposure is important for cosmetics.
7. For the generation of most kinetic data, non-animal methods are available or at an advanced stage of development. Given best working conditions, including resources in money and in manpower, alternative methods to predict renal and biliary excretion, as well as absorption in the lungs, need at least 5–7 years of development. However, the development of an integrated approach linking the results from in vitro/in silico methods with toxicokinetics modelling towards the full replacement of animals will take even more time.
8. However, it cannot be excluded that with the use of new exposure-driven risk assessment approaches, such as the TTC (threshold of toxicological concern), the need to replace at least some steps may become less relevant for regulatory decisions.

Objectives

Given the scenario of non-in vivo animal testing (1R) which has to be envisaged to be in place from 2013 on, the risk assessment of cosmetics is faced with a radically altered situation as compared with the 2005 report (see Coecke et al. 2005 in Eskes and Zuang 2005). In this new

⁸ 1R is the abbreviation for replacement (non-animal).

framework, exposure assessment is the important first step to decide on the necessity of further testing. Only in cases where a cosmetic ingredient is bioavailable following dermal, oral or inhalation exposure further tests on systemic and not just local toxicity would be necessary. The extent of exposure is compared with a dose which has a low probability to exert a toxic effect. This dose—also referred to as the threshold of toxicological concern (TTC)—is derived from the existing knowledge and could be used for chemicals for which little or no toxicological data are available.

Toxicokinetics is characterising the absorption, distribution, metabolism and excretion of a compound (ADME). ADME and biotransformation or metabolism encompasses all aspects of a pharmacokinetic/toxicokinetic evaluation. Studies to characterise steps in the toxicokinetic processes provide information about metabolite formation, metabolic induction/inhibition and other information which might be helpful for the study design of the downstream toxicological tests (the so-called toxico-dynamics). Metabolite/toxicokinetic data may also contribute to explaining possible toxicities and modes of action and their relation to dose level and route of exposure. Physiologically based toxicokinetic (PBTK)⁹ models are important to integrate the processes of absorption, distribution, metabolism and excretion (ADME) and are the tools to convert external exposure doses into internal concentrations and vice versa, thus enabling also for converting *in vitro* concentration–response into *in vivo* dose–response relationships. This chapter will illustrate that toxicokinetic data form a prerequisite for the conduct of other toxicological tests and are necessary to understand and interpret toxicological data. They are essential to extrapolate *in vitro* data to the human *in vivo* situation for the respective relevant toxicological endpoints.

It is seen necessary to maintain a dialogue between experts working on toxicokinetics and on toxicodynamics to ensure the interaction between the toxicokinetic and toxicodynamic processes are understood. The toxicity endpoints covered in other chapters deal with repeated dose exposures to xenobiotics, assessing chronic toxicities including target organ toxicities and target system toxicities, carcinogenicity, reproductive and developmental toxicity and sensitisation. It is necessary for toxicodynamic testing to take into consideration toxicokinetic processes of importance for the proper design and performance of *in vitro* toxicodynamic studies. Apart from ADME processes and their integration in PBTK models, *in vitro* biokinetics

measurements are further elements which characterise the concentration–time course during *in vitro* toxicity testing relevant for the concentration–effect relationship, the so-called actual concentration.

Background

The TTC concept

The TTC concept is an approach that aims to establish a human exposure threshold value below which there is a very low probability of an appreciable risk to human health, applicable to chemicals for which toxicological data are not available, based on chemical structure and toxicity data of structurally related chemicals.¹⁰ The TTC concept is currently used in relation to oral exposure to food contact materials, to food flavourings and to genotoxic impurities in pharmaceuticals and to metabolites of plant protection products in ground water (e.g. Kroes et al. 2004; Barlow 2005). Recently, The European Cosmetic Toiletry and Perfumery Association (COLIPA) sponsored work by a group of experts to examine the potential use of the TTC concept in the safety evaluation of cosmetic ingredients (Kroes et al. 2007). As the application of the TTC principle strongly depends on the quality, completeness and relevance of the databases, which are mostly based on toxicity data after oral exposure, its applicability to the dermal or inhalation uptake routes is limited, although it has been reported that the oral TTC values could be of some use for dermal exposures (Kroes et al. 2007); further, the use of TTC for the inhalation uptake route has recently been published (Escher et al. 2010; Westmoreland et al. 2010). However, only systemic effects are considered in the databases, and no local toxicity can be evaluated with this approach. Therefore, an improvement of the currently available databases is certainly needed. In addition, the TTC principle requires sound and reliable data on exposure (which might not always be available for cosmetics such as complex plant-derived mixtures) and the possibility to apply this principle in the field of cosmetics is still an ongoing discussion at an international level. Currently, the three Scientific Committees of DG SANCO have received a mandate to prepare an opinion on this topic.

If external exposure is above the external TTC, the toxicokinetic behaviour of a substance becomes important because it provides relevant information to derive the internal concentration related to the external exposure. The internal concentration, which might be different for different target organs, constitutes the basis for deciding on the necessity to perform further toxicity studies. The

⁹ In toxicological context, and also in the current legislation, the term physiologically based toxicokinetics (PBTK) is used, but a companion term physiologically based pharmacokinetics (PBPK) is used interchangeably especially in pharmaceutical contexts. There is no “deeper” difference between the two terms.

¹⁰ It is worth of noting, that some groups of substances are a priori exempted from the TTC approach (e.g. metals).

decision is made by comparing the internal concentration with a concentration which has a low probability to exert a toxic effect at organism level—the internal threshold of toxicological concern (TTC_{int})—derived from existing knowledge. The usefulness of the TTC_{int} concept is not yet widely discussed, but the concept is under development.

In vitro toxicokinetics as a key for 1R replacement strategies

Given the scenario of the development of compounds/products based on non-in vivo animal testing strategies (1R), toxicokinetics provides essential data for (1) establishing tools for PBTK modelling, (2) designing tests for toxicodynamic endpoints and (3) permitting a proper risk assessment. The implication for the 1R replacement paradigm is that toxicokinetic data would become the first data set to be produced using alternative methods.

Information on toxicokinetics under 1R is essential to address three major issues:

- (A) *Development and design of more efficient testing strategies:* As a key starting point for any toxicological testing, it is essential to know whether a substance will be bioavailable by one of the relevant uptake routes: only in cases where a cosmetic ingredient is bioavailable following dermal, oral or inhalation exposure, further tests on systemic and not just local toxicity will be necessary.
- (B) *In vitro–in vivo extrapolation:* To relate toxicodynamic information from non-animal-testing (1R) to real-life situation relevant for humans, i.e. to transform in vitro concentration–effect relationship into an in vivo dose–effect relationship. The most sophisticated challenge under 1R is to make in vitro data (from any type of toxicological endpoint) usable for risk assessment, i.e. to properly relate toxicodynamic information from in vitro studies to the in vivo situation, because test results under 1R will be presented as an in vitro concentration–effect relationship instead of an in vivo dose–effect relationship.
- (C) *Identification of clearance rates and the role of metabolites:* For the in vitro dynamics experiments, it is essential to know whether the cell or tissues are exposed to the parent compound and/or its metabolites. This information has to be known upfront based on toxicokinetic alternative methods identifying the main metabolites and the clearance rates of the parent compound and/or its metabolites.

Furthermore, nominal applied concentrations in in vitro media may greatly differ from the actual

intracellular concentration due to altered bioavailability (interactions with the medium, the plate, the cell itself) or to physiological cellular processes (mechanism of transport across the membranes, biotransformation, bioaccumulation). In repeated treatments for prolonged times of exposure, to mimic exposure to cosmetic products, the uncertainty about the actual level of exposure of cells in vitro is greatly enhanced. For this reason, in vitro biokinetics should be also considered in the experimental design for the in vitro dynamics experiments in order to correlate in vitro results to in vivo actual situations.

As amply justified earlier, it is without question that under 1R scenario of full animal replacement, toxicokinetic studies have to be performed differently from the study design and execution described in the OECD test guideline (OECD 417, the version adopted 1984; effective until September 2010). The newly effective and updated OECD 417 also comprises in vitro (e.g. use of microsomal fractions to address metabolism) and in silico (toxicokinetic modelling for the prediction of systemic exposure and internal tissue dose) methods (OECD 2010a). Essentially, instead of in vivo experiments, in vitro/in silico methods have to be used to derive the relevant information. This paradigm shift is illustrated schematically in Figs. 1, 2a and b.

The relation between kinetics and dynamics for a 1R replacement strategy

Ideally, and as a general goal, predictions of tissue exposure and subsequently toxicities should be based on human in vitro/in silico data combined with proper physiologically based toxicokinetic modelling, thereby replacing animal experiments. However, there are two questions that must be resolved in order to make the in vitro results usable for risk assessment. Firstly, the relationship between the effect of the parent compound and/or the metabolites on the in vitro test system and the health effect of interest must be clearly defined in order to derive a relevant in vitro (no)effect concentration or, better, benchmark concentration (BMC). In this respect, the actual rather than the nominal in vitro concentration tested, as stated previously, is a crucial starting point, to derive relevant parameters. Already at this stage, it is essential to consider what compound [e.g. the parent compound and/or metabolite(s)] the cells, tissues or organs will be exposed to, e.g. these data would be obtained by the kineticists. Secondly, an additional task of kineticists is to convert in vitro BMC to a predicted in vivo benchmark dose (BMD). For risk assessment purposes, the predicted in vivo BMD is to be compared to human exposure data (Rotroff et al. 2010).

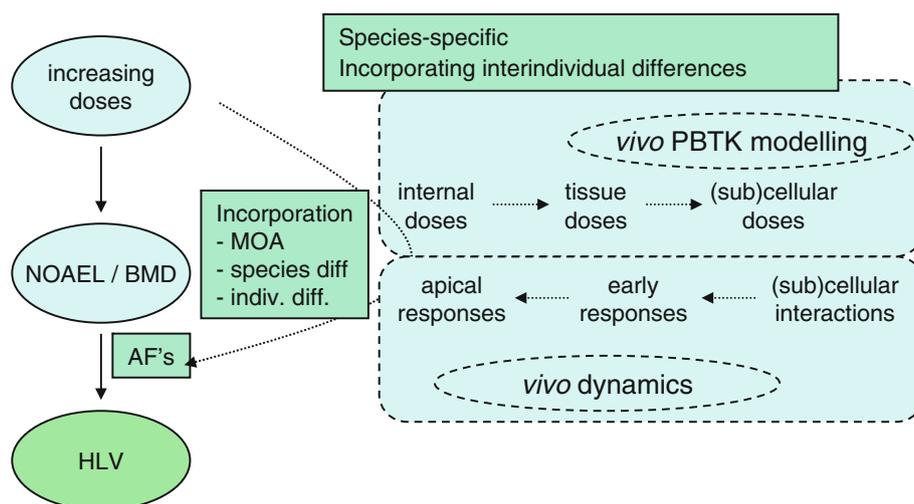


Fig. 1 Conventional human risk assessment (based on in vivo animal bioassays). *Solid ellipses on the left*: usually, animals (blue) are exposed at increasing doses, to derive an animal NOAEL or a BMD (POD; blue) which is converted to a human limit value (HLV) using appropriate assessment factors (AF's). Only the HLV is 'human' (green). *Dotted boxes on the right*: sometimes, AF derivation is based on further information from animal experiments such as on

mode of action, or possible species specificity. Definitions to this and the two following figures are in the footnote. *Definitions*: NOAEL No observed adverse effect level, BMD in vivo benchmark dose, BMC in vitro benchmark concentration, HLV human limit value (ADI, MAC, TLV, etc.); MOA mode of action, PBTK modelling physiologically based toxicokinetic modelling, AF assessment factor; *C,t* concentration–time

Importance of analytical methods in the 1R scenario

It is obvious that in this “alternative” scenario (1R), concentration measurements of the parent compound and/or the metabolites in the in vitro test system and the behaviour of a studied ingredient in the test system (in vitro kinetics or biokinetics) in general become an important part of the test design (Pelkonen et al. 2008a).

- Hence, a sensitive, specific and validated analytical and quantitative method for a new substance and its potential metabolites (Tolonen et al. 2009; Pelkonen et al. 2009a) will be pre-requisite in gathering data for quantitative risk assessment.
- Measurement of the actual rather than the nominal or ‘applied’ in vitro concentration¹¹ in the media or in cells is fundamental to performing in vitro kinetic modelling and in vitro studies on metabolism, preferably in human-derived systems.
- In vitro studies on distribution between blood/plasma and different tissues, in vitro absorption (gut, skin, lung) as well as studies on protein binding rely on the availability of appropriate analytical methods, although for this purpose in silico methods may also be available in the future, once the database containing chemical-specific toxicokinetic parameters evolves to an extent

that QSAR models can be built based on these parameters.

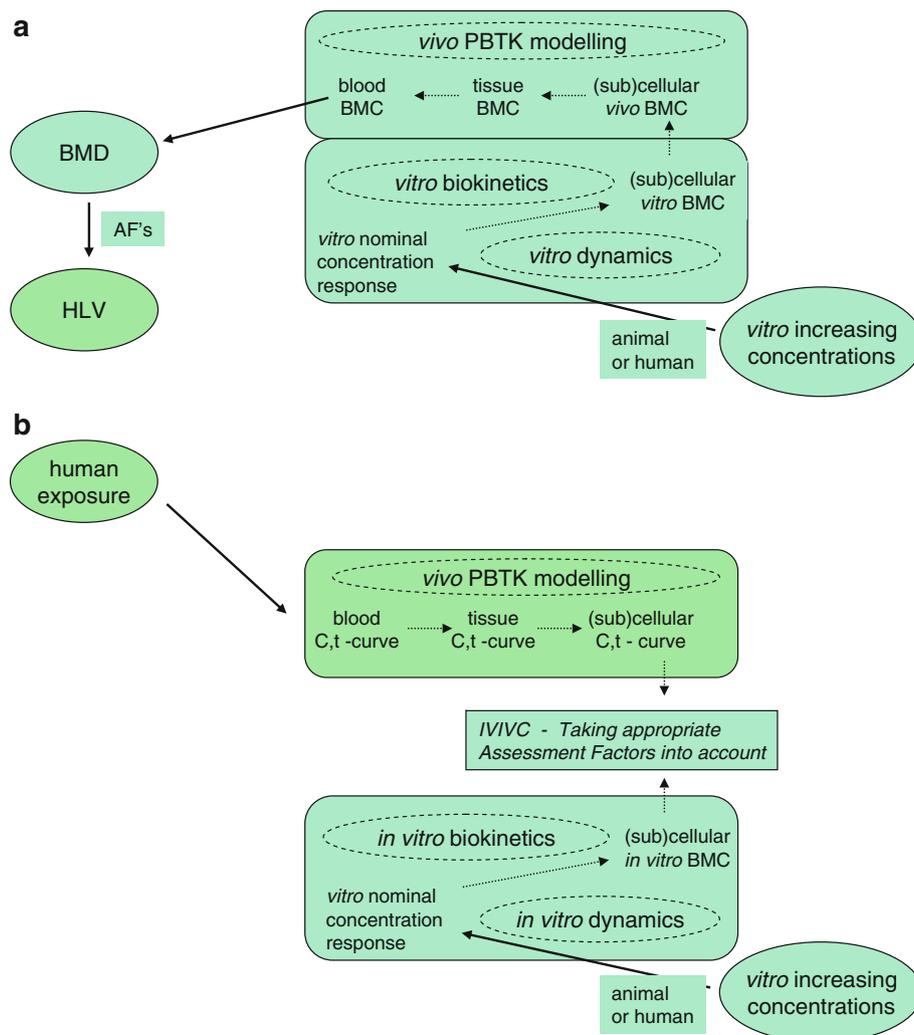
- Other fields of application for an analytical method are experiments to derive physico-chemical data, which are important as an input into QSAR for predicting the fate of substances.
- Chemical-specific measurements are also important as inputs into tissue composition-dependent algorithms to estimate the partitioning of chemicals into tissues.

Importance of actual, rather than nominal concentration in the 1R scenario

Kinetics has often been evoked to explain the differences between in vivo toxicity and results obtained in vitro, limiting the possibility to use the in vitro–derived data for an in vitro–in vivo extrapolation in the risk assessment of a chemical (Pelkonen et al. 2008a). Nevertheless, surprisingly few studies have addressed the issue of in vitro kinetics (Blaauboer 2010). One of the major problems of in vitro methods is the difficulty in the extrapolation of the dose–response relationship of toxicity data obtained in vitro to the in vivo situation if the nominal concentration of a chemical applied to cells is the basis for that extrapolation. PBTK models could help, allowing estimating tissues concentrations starting from a specific exposure scenario, or vice versa, to calculate from an effective dose resulting in vivo, the concentration resulting in a toxicologically relevant effect in an in vitro system (Mielke et al. 2010).

¹¹ In general, doses and concentrations should be expressed as moles (per appropriate denominator). Only for certain (practical) purposes mass unit could be used.

Fig. 2 **a** Future human risk assessment (no animal bioassays) based on in vitro–in vivo extrapolation (IVIVE) providing a human limit value (HLV). In vitro systems are exposed at increasing doses, to derive a NOAEC on the basis of biokinetic information, which is extrapolated by means of PBTK model to provide an in vivo human limit value. In vitro method, PBTK model and BMD can be animal or human-based (*blue-green*). HLV = human (*green*). **b** Future human risk assessment (no animal bioassays) based on in vitro–in vivo comparison (IVIVC). Predicted/modelled in vivo human internal exposure is compared to in vitro–derived (*bottom-up*) human internal benchmark concentration (BMC). In vitro method can be based on animal or human biological material (*blue-green*)



In any case, to use the nominal concentration in the in vitro system is a bad predictor of the free concentration and therefore a prerequisite for these extrapolations is the knowledge of the *actual* concentrations of the chemical exerting a toxic effect in the in vitro system. The nominal concentration, even when applied as a single dose, can to a great extent deviate from the actual concentration of the chemical in the system over time, due to altered bioavailability (interactions with the medium, adsorption to the disposable plastics, binding to proteins, evaporation) or to physiological cellular processes (mechanism of transport across the membranes, bio-transformation, bioaccumulation). In repeated treatments for prolonged time of exposure, the uncertainty about the actual level of exposure of cells in vitro is greatly enhanced also due to the metabolic capacity of the in vitro system. These processes have been shown to influence the free concentration and thus the effect (Gülden et al. 2001; Gülden and Seibert 2003; Heringa, et al. 2004; Kramer et al. 2009), clearly indicating

the need to estimate or measure the free concentration in the medium or the actual concentration in the cells (Zaldívar et al. 2010). One technique used to measure the free concentration in the medium is the solid-phase micro-extraction (SPME), the application of which showed that for some compounds the free concentration could differ up to two orders of magnitude from the nominal concentration (Kramer 2010).

The identification of in vitro relevant kinetic parameters, the elaboration of a tiered strategy to measure/estimate the real exposure of cells to xenobiotics and/or their metabolites in in vitro systems as key elements for IVIV extrapolation are among the major aims of PredictIV, a EU funded project, particularly of WP3: Non-animal-based models for in vitro kinetics and human kinetic prediction (see the Project website: <http://www.predict-iv.toxi.uni-wuerzburg.de/>). This is the first attempt in an EU Project to combine biological effects (toxicodynamics) with toxicokinetics and modelling ensuring the generation of real

exposure data linked to effects. In strict cooperation with the work package (WP) of the project dealing with the identification of effects (dynamic), the studies have been designed to determine the no observed effect concentration (NOEC) in model systems based on human cells representative of *in vivo* target organs. Data obtained will be modelled, in close cooperation with WP partners dealing with *in silico* methods, by using advanced PBPK modelling, so that starting from the NOECs, it will be possible to extrapolate the corresponding *in vivo* dose.

This approach is in line with the recommendation coming from an ECVAM-sponsored workshop on *in vitro* kinetics held in ISPRA in 2007, stating: In biologically relevant *in vitro* systems, good experimental design should always consider the impact of relevant *in vitro* factors, in particular kinetic factors, on the results. In order to achieve this, close cooperation between experimenters, modellers, biostatisticians and analytical chemists is necessary, particularly beyond the stage of prototype development.

Strategic considerations of risk assessment of cosmetic ingredients

Kinetics and dynamics are inherently linked to each other in the non-animal testing era, maybe even more than under the current situation, when animal bioassays are still allowed for several toxicological endpoints. This is illustrated in Fig. 3 (Bessems 2009). The left column starts on top with exposure and ends at the bottom with target tissue dose/concentration. The second column starts with the estimated/predicted target tissue dose/concentration which would indicate the range of concentrations to be tested *in vitro* with as sensitive techniques as possible (including omics). If *in vitro* effects are not measurable that are

predicted using the left column, any *in vivo* effects are quite unlikely and possible health risks would not be indicated.

With respect to prevalidation and the time required to prevalidation, it is worthwhile to spend a few sentences on the actual need to perform standard validation processes, such as by ECVAM. It is not uncommon that companies perform in-house validation, using often (historical) animal *in vivo* data. If the companies have a standard operating procedure for ADME testing, there may be chances to accept these kinds of methods via an independent expert opinion consultation under confidentiality agreements. This might be an alternative way to safeguarding consumer safety, by circumventing long-term validation programmes. Alternatively, if companies would be willing to cooperate, these in-house methods might be provided to others, as e.g. within COLIPA, and have independent experts reviewing the performance of the method including public consultation in a kind of small-scale prevalidation which could be sufficient for the purpose of ADME testing. Important in this respect is the inclusion of well-known reference compounds. In addition, for ADME testing, very high precision of a method covering one of many aspects determining blood concentration–time curves might not always be key, especially not during pre-screening when validation of a PBPK modelling prediction would always be a case-by-case validation, not a method validation. Here, robustness may be a much more important criterion than accuracy.

Under the new, non-whole-animal testing paradigm, the first step of a screening process of chemicals for possible use as a cosmetic ingredient should be to find out whether absorption is likely or not under the foreseen use scenario. To this end, the following decision tree (Fig. 4) might be very helpful before starting any effects testing (both *in vivo* or *in vitro*). Ultimately, companies could decide to stop further development, if a compound appears to be absorbed under foreseen circumstances.

One step further, if a chemical appears to be absorbed to a limited extent, enough not to block further R&D, is to assess its systemic exposure quantitatively (Fig. 5). This will deliver essential information for the *in vitro* effects testing.

Available non-animal methods to derive values for absorption, distribution, metabolism and excretion

In this section, a general survey on the current status of non-animal methods for deriving input parameters for PBTK modelling is presented. For details regarding several mostly *in vitro* and *in silico* methods that are under development at various stages, the reader is referred to the Supplementary Online Information, which includes several tables.

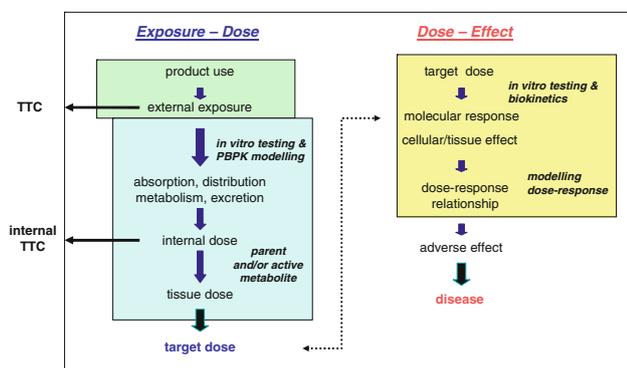
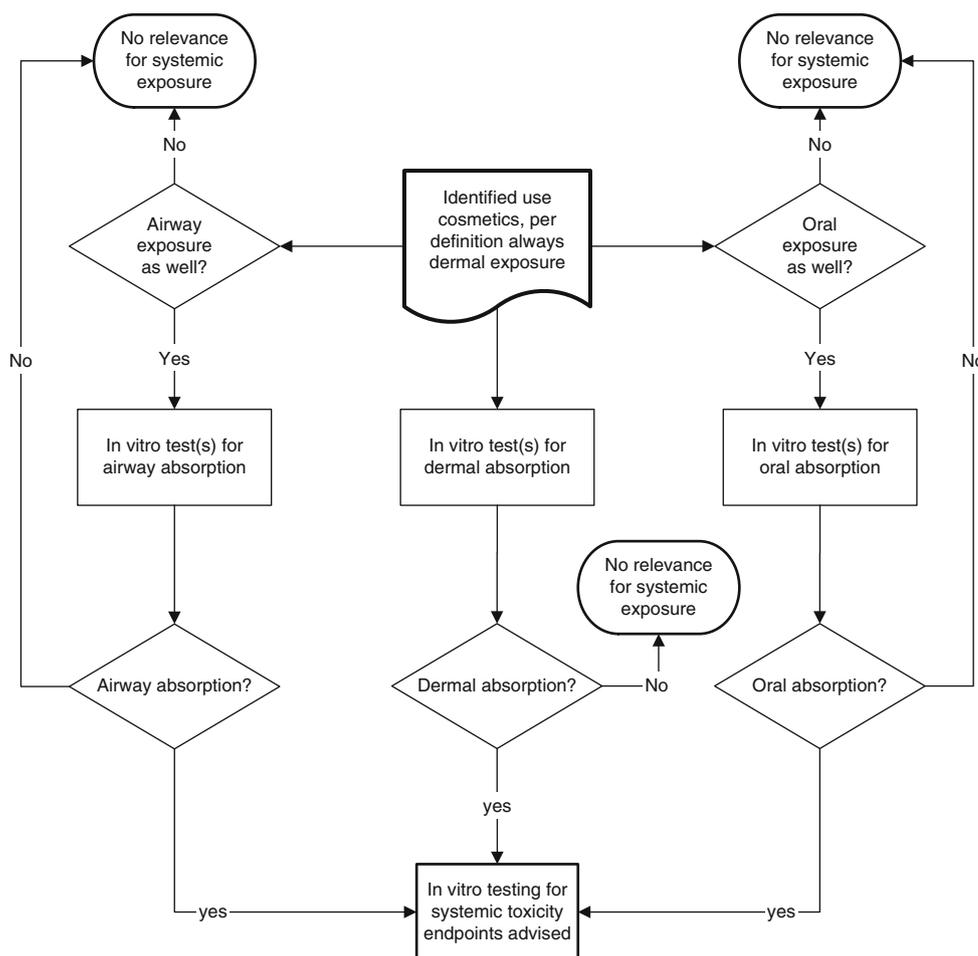


Fig. 3 The interdisciplinary link between kinetics and dynamics. From external exposure to possible *in vitro* indications of effects (amended from Bessems 2009)

Fig. 4 Decision tree for absorption-based testing

Absorption and bioavailability after dermal, inhalatory or oral exposure

Absorption is the transport across an epithelial layer. Bioavailability is more complex and is defined as the fraction of a chemical in a certain matrix that reaches the systemic circulation unchanged. In that way, it is a complex parameter, describing several processes. Although it is difficult to study in isolation in such a way that the outcome is easily applicable in PBTK models, it is important in risk assessment and possibly in intelligent testing strategies.

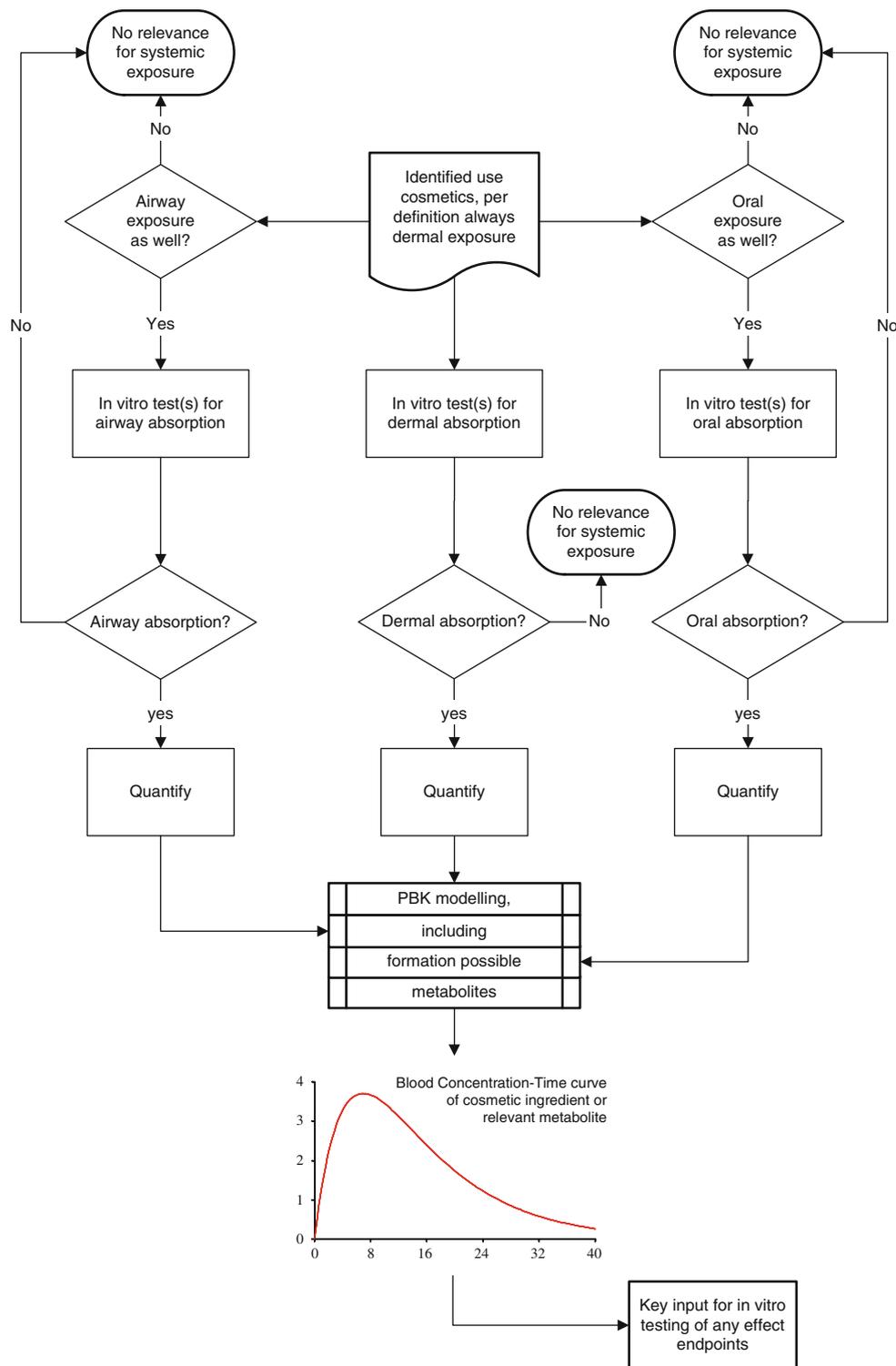
Bioaccessibility models Release of the compound from its matrix is required for transport across the dermal, lung or intestinal epithelium and bioavailability of a compound to the body. Release depends mainly on the matrix–compound interaction. In vitro models to assess bioaccessibility are best developed for the oral route (Brandon et al. 2006). Dermal exposure and exposure via inhalation are more problematic. Some methods for these types of exposure are

under development for other areas than cosmetics. They will have to be developed further for application to cosmetic ingredients in the next years.

Absorption models Absorption depends on the compound-specific properties and physiology (and pathology) of the epithelial tissue. Some of the properties of importance are physico-chemical properties of a compound and the availability of specific influx and efflux transporters in the tissue.

Dermal exposure: Various in silico QSAR models are available for the skin permeability prediction of compounds although none of them has been developed for a broad applicability, i.e. for a broad range of physico-chemical properties. Although some of them have been set up according to the OECD Principles for the validation of QSAR (OECD 2004a) and may be useful for specific chemicals, none of them has been widely validated up to our knowledge (Bouwman et al. 2008). Regarding in vitro absorption models, OECD Technical Guideline 428 is

Fig. 5 Decision tree for (internal) exposure-based testing



available (OECD 2004b) and guidance is presented as well by OECD (2004c), US EPA (2004), EFSA (2009) and SCCNFP (2003).

Inhalatory exposure: The lung can be anatomically divided into several parts: trachea, bronchi, bronchioles and alveoli. In the upper respiratory airways, the absorption

is low, and it mostly occurs in the lower part. No QSAR models predicting lung absorption are known in the public literature. In vitro models to study the translocation of compounds in the lung are in various stages of development. Several more years of intensive research will be needed to provide suitable models that can enter prevalidation.

Oral exposure: In silico QSAR-like models can predict specific parameters for an unknown chemical based on structural and physicochemical similarities to various known chemicals. In vitro models such as the Caco-2 cell line can predict the absorption over a single barrier as well and are rather standard. They could be incorporated in a medium-throughput test strategy. Importantly however, and as with in silico models, the validity of these in vitro model predictions for cosmetic ingredients remains to be established, because most, if not all, of these models were developed for pharmaceutically active ingredients. While in the pharmaceutical R&D reliable prediction between 50 and 100% absorption is important, in the cosmetics arena the crucial range will be the lower absorption range, i.e. much further less than 10% or even 1% absorption. Many more years of intensive research seem necessary for cosmetic ingredients before prevalidation of in vitro or other methods that are suitable to assess the potential pneumonal toxicity of cosmetic ingredients comes in sight.

Bioavailability Three processes (partially linear, partially in parallel) can be distinguished that determine bioavailability: (1) release of the compound from its matrix (bioaccessibility), (2) absorption of the released fraction and (3) metabolism before reaching the systemic circulation (Oomen et al. 2003).

In order to predict the precise bioavailability of a cosmetic ingredient from a cosmetic product, it is therefore important to determine the three different processes involved. However, the bioaccessibility could be used as a measure for the maximal bioavailability. If the parent compound would be expected to cause toxicity, complete absence of bioaccessibility would indicate absence of systemic effects. There are in vitro models to measure absorption through various portals of entry, although they are at various stages of development. In vitro models for measuring metabolism are described in section “[Metabolism \(biotransformation\)](#)”.

In silico models to estimate oral bioavailability (Bois et al. 2010) have been developed for the use in conjunction with PBTK models. The use of these organ-level in silico models is currently the best way to integrate the inputs from tests of bioaccessibility, absorption and metabolism, including hepatic clearance in the first-pass situation, because of the complex nature of bioavailability. However clearly, relevance and reliability of these in silico models

outside the pharmaceutical R&D will need quite some years of extra investigations before prevalidation, the capacity to demonstrate reliability, would be reachable.

Distribution

After absorption, the distribution of a compound and its metabolites inside the body is governed by three main factors: (1) the partition of the substance with plasma proteins, (2) between blood and specific tissues and (3) the permeability of the substance to cross specialised membranes, so-called barriers (e.g. blood–brain barrier/BBB, blood–placental barrier/BPB, blood–testis barrier/BTB).

Estimation of plasma protein binding (PPB) Only the free (unbound) fraction of a compound is available for diffusion and transport across cell membranes. Therefore, it is essential to determine the binding of a compound to plasma or serum proteins. The easy availability of human plasma has made it possible to determine the unbound fraction of compounds by performing in vitro incubations directly in human plasma.

In vitro approaches: There are three methods generally used for PPB determination: (1) equilibrium dialysis (ED) (Waters et al. 2008), (2) ultrafiltration (UF) (Zhang and Musson 2006) and (3) ultracentrifugation (Nakai et al. 2004). All methods can be automated for high throughput, are easy to perform and have good precision and reproducibility. The use of combined LC/MS/MS allows high selectivity and sensitivity. Equilibrium dialysis is regarded as the “gold standard” approach (Waters et al. 2008).

In silico approaches: Two recent reviews on the in silico approaches for the estimation of PPB have been carried out by Wang and Hou (2009) and Mostrag-Szlichtyng and Worth (2010). A general correlation based on the octanol–water partition coefficient was proposed by de Bruyn and Gobas (2007) after a compilation of literature data and using a broad variety of chemicals, i.e. pesticides, polar organics, polychlorinated biphenyls, dioxins, furans, etc.

Estimation of blood–tissue partitioning The fate of a compound in the body is determined by partitioning into the human tissues. Therefore, the knowledge of this partitioning is of fundamental importance for the understanding of a compound’s kinetic behaviour and toxic potential. The measurement of tissue storage and a molecular understanding of tissue affinity have, historically, not been studied to the same extent as plasma protein binding; however, the knowledge of these partitioning coefficients is essential for the development of PBTK models. Fortunately, quite a large number of approaches have been developed over the last years.

In vitro approaches: The available system is the vial-equilibration technique. A spiked sample of organ tissue–buffer homogenate is equilibrated and subsequently, the free (unbound) concentration of the test chemical is determined. The tissue–blood partition coefficient is calculated using results from pure buffer, tissue–buffer and blood–buffer incubations. Tissues can be mixed to obtain average values for example richly perfused tissue groups. Olive oil or octanol are often used instead of adipose tissue. The free (unbound) concentration is typically assessed by one of the following techniques: equilibrium dialysis, ultracentrifugation, headspace analysis (for volatiles) or solid-phase (micro-) extraction followed by a classical analysis such as HPLC UV or MS. The purpose of this technique is the prediction of the in vivo tissue blood partitioning and the prediction of an in vivo volume of distribution (Gargas et al. 1989; Artola-Garicano et al. 2000).

In silico approaches: Different methodologies have been developed, starting from QSAR, correlations with physicochemical properties, up to mechanistic approaches. The main problem for the generalisation of QSAR correlations has been the poor results obtained for charged molecules under physiological conditions and with charged phospholipids. However, there are mixed and mechanistic approaches with tolerable error ranges (Poulin and Theil 2002, 2009; Schmitt 2008).

Estimation of substance permeability through specialised barriers Blood–Brain Barrier (BBB): The BBB is a regulatory interface that separates the central nervous system (CNS) from systemic blood circulation and may limit or impair the delivery of certain compounds, which makes the brain different from other tissues. There are several passive and active mechanisms of transport through the BBB (Mehdipour and Hamidi 2009).

Several in vitro BBB models are under development which integrate various cell of vascular and neural origin. Also single cell lines containing transfected transporters have been proposed as models to study BBB permeability. However, all available models are in early stages of development. Several in silico models exist to predict BBB penetration, although the vast majority of these approaches do not consider transport mechanisms taking place (Mostrag-Szlichtyng and Worth 2010). Recently, some molecular models have been developed to consider the BBB transporters (Allen and Geldenhuis 2006).

Blood–Placenta Barrier (BPB): The BPB serves to transport nutrients and waste, and other compounds such as hormones. However, the placenta does not provide a true barrier protection to the foetus from exposure to compounds present in the mother's systemic circulation, although it might reduce the transport of certain molecules.

The transfer across the placenta can occur by several active or passive processes (Myren et al. 2007).

Experimental methods to study human transplacental exposure to toxic compounds have been reviewed by Vähäkangas and Myllynen (2006). There are both primary and permanent trophoblast-derived cell models available. An ex vivo model, human perfused placenta cotyledon, offers information about transplacental transfer, placental metabolism, storage, acute toxicity and the role of transporters, as well as an estimation of foetal exposure. There are also few QSAR models, although active transport mechanisms and potential metabolism are not addressed.

Blood–Testis Barrier (BTB): In the testis, the BTB is a physical and physiological barrier which assures functions in hormonal regulation and spermatogenesis (Fawcett et al. 1970). Many systems have been tested as organ cultures, co-cultures or single cell cultures, but none has really developed for considering toxicokinetic processes.

Metabolism (biotransformation)

Metabolism or biotransformation is the principal elimination route of organic chemicals; roughly 70–80% of pharmaceuticals are partially or practically completely eliminated by metabolism (Zanger et al. 2008). Due to a multitude of xenobiotic-metabolising enzymes possibly acting on a chemical with different metabolic pathways, the first screen should preferably be as comprehensive as possible. Because liver is the principal site of xenobiotic metabolism, the enzyme component in in vitro systems should preferably be liver-derived (Coecke et al. 2006; Pelkonen et al. 2005, 2008b) and of human origin, to avoid species differences (see e.g. Turpeinen et al. 2007). There is a generally accepted consensus that metabolically competent human hepatocytes or hepatocyte-like cell lines are the best enzyme source to perform the first primary screening of metabolism (Gómez-Lechón et al. 2003, 2008; Houston and Galetin 2008; Riley and Kenna 2004). The two most important endpoints measured are (1) intrinsic clearance which can be extrapolated into hepatic metabolic clearance and (2) the identification of metabolites (stable, inactive, active or reactive metabolites of concern).

In silico approaches The available systems are of three types: (1) expert systems based on structure-metabolism rules, (2) SAR and QSAR modelling and (3) systems based on pharmacophore (ligand) or target protein modelling. There are a large number of commercial softwares available for predicting biotransformation, in various phases of development. Although in silico approaches are developing rapidly, they are still inadequate for the production of results which are accepted by regulators, and new approaches are needed to predict the major metabolic

routes when there are a number of potential metabolic pathways. Reliable and good-quality databases (not limited to pharmaceuticals) are of the utmost importance for the development of reliable software for application to a wider assortment of chemicals including cosmetic ingredients and they are still in great need. Discussions of various approaches can be found in recent reviews (Testa et al. 2004; de Graaf et al. 2005; Kulkarni et al. 2005; Crivori and Poggesi 2006; Lewis and Ito 2008; Muster et al. 2008; Mostrag-Szlichtyng and Worth 2010).

Metabolic clearance The metabolic stability test is a relatively simple, fast-to-perform, but specialised analytical equipment-based MS study, to find out whether a compound is metabolically stable or labile. It is based on the disappearance of the parent compound over time (with the appropriate analytical technique) when incubated with a metabolically competent tissue preparation (e.g. a human liver preparation, preferably human hepatocytes). The rate of parent compound disappearance gives a measure of its metabolic stability and allows for the calculation of intrinsic clearance and extrapolation to hepatic (metabolic) clearance. The use of liver-based experimental systems should give a fairly reliable view of hepatic intrinsic clearance. However, to be able to predict *in vivo* clearance, a number of assumptions concerning the substance under study must be made, so an extrapolation model is needed (see e.g. Pelkonen and Turpeinen 2007; Rostami-Hodjegan and Tucker 2007; Webborn et al. 2007). Although no formal validation studies are known, the screening test for metabolic clearance should be relatively ready for validation after the availability of common procedures and related SOPs.

To cover also extrahepatic biotransformation, the above-described method for metabolic stability can be combined with the use of other tissues. For cosmetic substances, dermal uptake is the most prominent intake pathway and consequently methodologies for skin metabolism would be of considerable significance. Likewise, inhalation (spray applications) is also an important uptake route, and metabolism should be taken into consideration in *in vitro* pulmonary tests. Some efforts would be needed to standardise metabolic stability in skin and pulmonary tissues.

Metabolite profile and bioactivation With the advent of modern MS techniques, it is possible and feasible to study both the detailed qualitative and quantitative metabolic profiles of a compound (Pelkonen et al. 2009b; Tolonen et al. 2009). The use of recombinant enzymes, transfected cells and metabolically competent human (liver-derived) cell lines or subcellular fractions is being very actively employed in pharmaceutical industry and academia. In this way, it is possible to have indication about the enzymes

participating in the metabolism which allows a number of predictions about physiological, pathological and environmental factors affecting the kinetics of a compound of interest.

The formation of reactive metabolites by biotransformation seems to be the cause of deleterious effects for a large number of compounds (Park et al. 2006; Williams 2006). Even though mechanistic details of relationships between toxicities and reactive metabolites are still somewhat unclear, there is ample indirect evidence for their associations (Baillie 2006, 2008; Tang and Lu 2010).

There are direct and indirect methods to test the potential formation of reactive metabolites. Most direct assays use trapping agents (i.e. glutathione or its derivatives, semicarbazide, methoxylamine or potassium cyanide) that are able to trap both soft and hard electrophiles: conjugates are then analytically measured. The Ames test is a prime example of an indirect method for the bioactivation assay making use of metabolically competent enzyme system (which could be human-derived, if needed) and properly engineered bacteria to detect reactive, DNA-bound metabolites.

Induction assays Since induction has a complex underlying mechanism, it is a good indicator for high-quality metabolic competent systems that can be used for long-term purposes (Coecke et al. 1999; Pelkonen et al. 2008b): that is why developments are ongoing to assess CYP induction in bioreactor-based systems. Obviously, the most relevant intake routes for cosmetics (dermal, inhalation) should be considered when *in vitro* test systems are developed.

A large number of test systems ranging from nuclear receptor binding assays to induction-competent cell lines and cryopreserved human hepatocytes are currently available. The reliability of 2 hepatic metabolically competent test systems, e.g. cryopreserved hepatocytes and cryopreserved HepaRG systems, is currently assessed by ECVAM (International Validation Trial) by using CYP induction at the enzyme level as the endpoint detection method. These test systems are widely used in pharmaceutical industry to help early drug development and are designed to detect induction of CYP enzymes relevant for the pharmaceutical area. This can represent a potential limitation since, for cosmetics, other CYP forms might play an additional or more prominent role. Thus, further progress is needed to cover this potential gap.

Inhibition assays Due to the broad substrate specificity of metabolising enzymes, there is always a possibility that compounds would interfere with each other's biotransformation. Inhibition of biotransformation leads to higher concentrations and delayed clearance and may cause adverse effects. At the site of entry (i.e. GI tract, skin,

lung), inhibition of the first-pass metabolism would increase the blood concentration of the parent compound.

There are currently available a large number of test systems ranging from recombinant expressed enzymes (principally CYP and UGT enzymes, but increasingly also other xenobiotic-metabolising enzymes) to primary cells (hepatocytes) and permanent cell lines (Li 2008; Farkas et al. 2008). All these test systems are widely used in pharmaceutical industry and can be judged to be validated at least for pharmaceuticals. This can represent a potential limitation since, for cosmetics, other CYP forms might be relevant. In addition, some cosmetics contain complex plant-derived mixtures, and it is not elucidated to what extent current inhibition assays would be applicable. Thus, further progress should cover chemical and compositional peculiarities characteristic for the cosmetics field.

Excretion

Predicting major excretion pathways of compounds is important in relation to their kinetic behaviour and the relationship to pharmacological/toxicological effects. The kidneys and the hepatobiliary system have the capacity to excrete either as the parent compound or as metabolites and are important routes for elimination of xenobiotics and their metabolites. Unfortunately, excretory processes seem to be the least developed area in the context of *in vitro* toxicokinetic methods probably because renal and biliary excretion, the major excretory routes, are complex processes with a number of passive components and active processes involved.

Renal excretion: Excretion by the kidney encompasses three different mechanisms and they all include the interplay of both passive movement of drugs and the participation of a number of active transporters. Even if there are examples how the involved transporters can be identified, it is difficult to use the findings to feed into a physiological model of renal excretion which includes tubular secretion and tubular reabsorption.

Biliary excretion: In humans, biliary excretion does not seem to play an important role for most of the substances. However, in cases where it matters, the process is rather complex, first preceded with the entry of the substance to the hepatocyte and its possible metabolism by the hepatic metabolic machinery. With most substances ultimately excreted into the bile, phase II metabolising enzymes produce conjugates, which are then transported across the canalicular membrane to be excreted into the bile.

Current approaches and future efforts needed. There have been few attempts for developing expert systems or computational approaches to predict renal excretion from some basic molecular and physicochemical properties. Likewise, *in silico* modelling attempts are being made to evaluate the molecular weight dependence of biliary excretion (well established in rats) and to develop

quantitative structure–pharmacokinetic relationships to predict biliary excretion. Efforts have been undertaken to use collagen-sandwich cultures of hepatocytes as an *in vitro* test system for testing biliary excretion. Due to the fact that progress in the field is very recent, no systematic efforts have been undertaken to standardise the above mentioned approaches. Some pharmaceutical companies as well as academic groups have published reports on their experiences. No formal validation studies are known.

An understanding of mechanisms that determine these processes is required for the prediction of renal and biliary excretion. Physiologically based *in vitro/in-silico/in vivo* approaches could potentially be useful for predicting renal and biliary clearance. Whereas for biliary excretion some advances have been made with *in vitro* models (i.e. sandwich-cultured hepatocytes), no reports could be identified in the literature on *in vitro* models of renal excretion nor were reports available on *in silico* methods.

Integrating *in vitro* and *in silico* approaches using PBTK modelling

After a chemical compound penetrates into a living mammalian organism (following intentional administration or unintentional exposure), it is usually distributed to various tissues and organs by blood flow (Nestorov 2007). The substance can then bind to various receptors or target molecules, undergo metabolism or can be eliminated unchanged. The four processes of absorption, distribution, metabolism and elimination (ADME) constitute the pharmacokinetics of the substance studied. The term toxicokinetics is used if the substance is considered from a toxicity view point.

In general, the toxicokinetics of a compound are the function of two sets of determinants: physiological characteristics of the body (which are compound independent) and compound-specific properties. It is possible to quantify some compound-specific structural properties and to relate them to biological activity. That is the basis of the so-called Quantitative Structure–Activity Relationships (QSARs). Likewise, structural properties can be used to estimate other properties, such as lipophilicity ($\log K_{ow}$) and blood over tissue partition coefficients. In that case, the term Quantitative Structure–Property Relationship (QSPR) is used. To quantitatively predict the toxicokinetics of a substance, it is necessary to model jointly their physiological determinants and the compound-specific properties. Physiologically based toxicokinetic (PBTK)¹² modelling is currently the most advanced tool for that task.

¹² In practically all sources on which this section is based, the term physiologically based pharmacokinetics (PBPK) is used, but for the sake of subject matter and chemicals legislation, PBTK is used throughout in this report.

PBTK models are necessary tools to integrate in vitro and in silico study results

The concentration versus time profiles of a xenobiotic in tissues, or the amount of its metabolites formed, is often used as surrogate markers of internal dose or biological activity (Andersen 1995). When in vivo studies cannot be performed or when inadequate in vivo data are available, the toxicokinetics of a substance can be predicted on the basis of in vitro or in silico studies. For risk assessment purposes, in vitro systems should be mechanism based and able to generate dose/concentration–response data. The greatest obstacles to the use of in vitro systems are the integration of their data into a biologically meaningful framework and their extrapolation to in vivo conditions. PBTK models are ideally suited for this, because they can predict the *biologically effective* dose of an administered chemical at the target organ, tissue and even cell level (Barratt et al. 1995; Blaauboer et al. 1999; Blaauboer et al. 1996; Combes et al. 2006; DeJongh et al. 1999a; Dr. Hadwen Trust Science Review 2006). Indeed, PBTK models are increasingly used in drug development and regulatory toxicology to simulate the kinetics and metabolism of substances for a more data informed, biologically based and quantitative risk assessment (Barton et al. 2007; Boobis et al. 2008; Bouvier d’Yvoire et al. 2007; Loizou et al. 2008; Meek 2004). As such, they should be able to significantly reduce or even replace animals in many research and toxicity studies.

General description of PBTK models

A PBTK model is a mechanistic ADME model, comprising compartments that correspond directly to the organs and tissues of the body (e.g. liver, lung, muscle), connected by the cardiovascular system (see Fig. 6). The main application of PBTK models is the prediction of an appropriate target tissue dose, for the parent chemical or its active metabolites. Using an appropriate dose-metric provides a better basis for risk assessment (Barton 2009; Conolly and Butterworth 1995). The estimation of dose-metrics is regarded as the ‘linchpin’ of quantitative risk assessment (Yang et al. 1998). In the IR approach, the question may be to predict the external dose leading to an internal dose equivalent to that of a given in vitro treatment.

PBTK models’ parameter values can be determined on the basis of:

- in vitro data,
- in vivo data in humans,
- quantitative structure–property relationship (QSPR) models,
- the scientific literature.

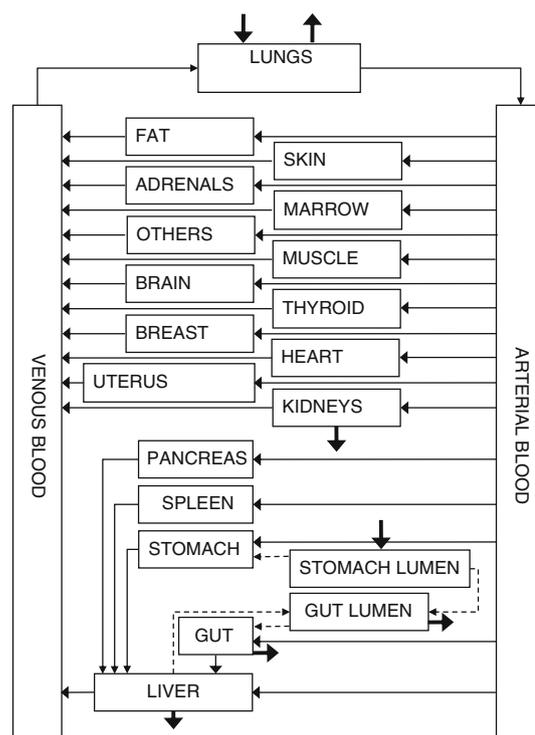


Fig. 6 Schematic representation of a PBTK model (for a woman). The various organs or tissues are linked by blood flow. In this model, exposure can be through the skin, the lung or *per os*. Elimination occurs through the kidney, the GI tract, and the lung, or by metabolism in the liver. The parameters involved are compartment volumes, blood flows, tissue affinity constants (or partition coefficients), and specific absorption, diffusion, metabolic, and excretion rate constants. The whole life of the person can be described, with time-varying parameters. The model structure is not specific of a particular chemical (see <http://www.gnu.org/software/mcsim/>, also for a pregnant woman model)

Published models range from simple compartmental (Gibaldi and Perrier 1982; see also Pelkonen and Turpeinen 2007 for current practical solutions) to very sophisticated types (Jamei et al. 2009). Between compartments, the transport of substances is dictated by various physiological flows (blood, bile, pulmonary ventilation, etc.) or by diffusions (Gerlowski and Jain 1983; Bois and Paxman 1992). Perfusion-rate-limited kinetics applies when the tissue membrane presents no barrier to distribution. Generally, this condition is likely to be met by small lipophilic substances. In contrast, permeability-rate kinetics applies when the distribution of the substance to a tissue is rate-limited by the permeability of a compound across the tissue membrane. That condition is more common with polar compounds and large molecular structures. Consequently, PBTK models may exhibit different degrees of complexity. In the simplest and most commonly applied form (Fig. 6), each tissue is considered to be a well-stirred compartment, in which the substance distribution is limited by blood

flow. In such a model, any of the tissues can be a site of elimination. However, in Fig. 6, it is assumed that the liver is the only metabolising organ and that excretion only happens in the kidney.

Building a PBTK model requires gathering a considerable amount of data which can be categorised into three groups: (1) system data (physiological, anatomical, biochemical data), (2) compound-specific data and (3) the model structure, which refers to the arrangement of tissues and organs included in the model (Rowland et al. 2004). In a sense, PBTK modelling is an integrated systems approach to both understanding the kinetic behaviour of compounds and predicting concentration–time profiles in plasma and tissues. Additional details of PBTK modelling and applications can be found elsewhere (Gerlowski and Jain 1983; Nestorov 2003; Rowland et al. 2004; Jones et al. 2009; Edginton et al. 2008; Pelkonen et al. 2008a; Kapitulnik et al. 2009; Dahl et al. 2010). Indeed, such descriptions of the body are approximate, if not rough, but a balance has to be found between precision (which implies complexity) and simplicity (for ease of use). Yet, the generic structure of a PBTK model facilitates its application to any mammalian species as long as the related system data are used. Therefore, the same structural model can approximately be used for a human, a rat or a mouse (De Buck et al. 2007).

Generic applications of PBTK modelling

Inter-individual or intra-individual extrapolations: These refer to the fact that a given exposure may induce different effects in the individuals of a population and that the same individual may respond differently to the same exposure at different times in his/her lifetime. These extrapolations are performed by setting parameter values to those of the subpopulation or individual of interest and are mainly used to predict the differential effects of chemicals on sensitive populations such as children, pregnant women, the elderly, the obese, and the sick, taking into account genetic variation of key metabolic enzymes, etc. (Jamei et al. 2009). The toxicokinetic behaviour of a compound can also be studied under special conditions, such as physical activity.

Inter-dose extrapolations: These extrapolations are achieved by capturing both the linear and non-linear steps of the biological processes known to govern the kinetics of the chemical of interest, e.g. in the transport and metabolism.

Inter-route of exposure extrapolations: Any route of exposure can be described either in isolation or in combination. For example, systemic toxicity may be studied following intravenous infusion, uptake via the gastrointestinal tract, dermal absorption and inhalation via the lungs. For example, coumarin hepatotoxicity is dependent on the route of administration and can be rationalised on

the basis of physiologically based modelling (Kapitulnik et al. 2009).

Specific applications of PBTK modelling in the case of the IR for cosmetics

A tiered approach for pure predictions of toxicity: PBTK models can be used in a step by step or tiered approach. They can be first coupled to *in silico* quantitative structure–pharmacokinetics properties relationships (QSPR) models for partition coefficients, absorption or excretion rate constants and computer models of metabolism (assuming that such models are available for the chemical class of interest). Using expected exposure patterns, estimates of internal exposures, bioavailability, half-life, etc. can be obtained. Such results could either be sufficient to answer the question of interest or would provide at least estimates of concentration levels to be assayed *in vitro*. In further steps, leading to increased refinement and predictive accuracy, PBTK models can incorporate the results of specific *in vitro* estimates of pharmacokinetic parameters (such as absorption rates, metabolic rate constants, etc.). At any point of that approach, the PBTK model provides estimates of internal dose levels attained in predefined exposures scenarios, enabling a prediction of the most sensitive toxic endpoint, of exposure–response relationships, of no-effect levels (if the dynamic models provide toxicity thresholds), etc. (Fig. 7).

Forward dosimetry: *in vitro*–*in vivo* correlation: Historically, in chemical risk assessment, PBTK modelling has been used primarily for ‘forward dosimetry’, that is, the estimation of internal exposures in the studies characterising the toxicity of a chemical. The human chemical risk

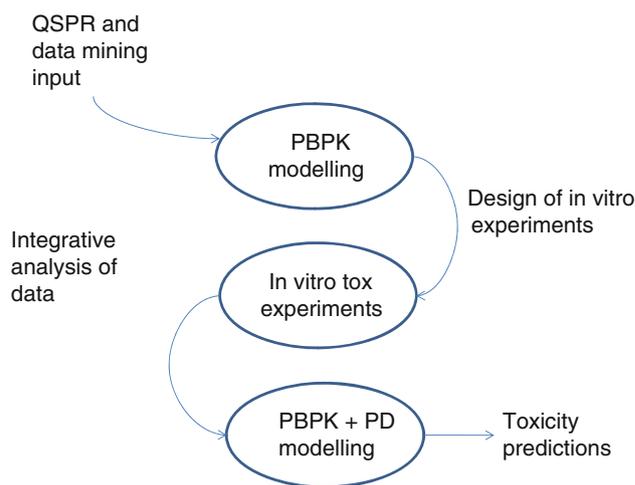


Fig. 7 A tiered approach for experimental design and predictive toxicity assessment using PBTK and pharmacodynamic (PD) modelling

assessment arena may be described as ‘data poor’ as opposed to the ‘data-rich’ pharmaceutical arena, hence, the need to estimate internal exposure through modelling, in the absence of specific measurements. When *in vitro* systems, such as human cell lines, will replace animals in toxicological and safety evaluation of cosmetics, we will also need to estimate *in vivo* internal doses. This will require PBTK modelling, as illustrated in Fig. 7.

Reverse dosimetry: exposure reconstruction from *in vitro* alternatives: Recently, a number of studies have attempted to ‘reconstruct dose’ or ‘estimate external exposure’ consistent with human biological monitoring data. That exercise has been described as ‘reverse dosimetry’ (Clewel et al. 2008; Georgopoulos et al. 1994; Liao et al. 2007; Lyons et al. 2008; Roy and Georgopoulos 1998; Tan et al. 2006a, b). A similar procedure could be applied to estimate the external exposure levels leading to acute and chronic systemic toxicity, including repeated dose systemic toxicity, as estimated from *in vitro* alternative methods.

Exposure reconstruction can, and should, be addressed at both the individual and population level. Population-based estimates of exposure should account for human inter-individual variability, both in the modelling of chemical disposition in the body and in the description of plausible exposure conditions.

The reconstruction of dose or exposure using Bayesian inference is recommended, even for systems where tissue dose is not linearly related to external exposure (Allen et al. 2007; Lyons et al. 2008; Sohn et al. 2004). Gelman et al. (1996) presented a general method of parameter estimation in PBTK models, and reverse dosimetry is a type of PBTK model calibration problem.

Current limitations: All the limitations of *in vitro* toxicokinetic assays have an impact on the predictive accuracy of PBTK models. Difficulties in predicting metabolism, renal excretion and active transport are foremost in that respect, and improvements will proceed at the pace adopted to solve these problems. More intrinsic to PBTK modelling itself is the difficulty to accurately model dermal exposure (e.g. surface area exposed, dose applied wearing and washout) and absorption (e.g. saturation of the skin layers), at least for some important chemical classes (PCBs, etc.). The most precise solutions involve partial differential equation models, even though various approximations are available (Krüse et al. 2007). This goes beyond the capabilities of commonly used PBTK modelling software, and a particular effort would need to be devoted to resolving that problem.

Checking the validity of PBTK models is much easier when they have a stable and well-documented physiological structure. That is a particular advantage of the generic PBTK models developed by Simcyp (<http://www.simcyp.com>),

Bayer Technology Services (<http://www.pk-sim.com>), Cyprotex (<https://www.cloegateway.com>) or Simulation Plus (<http://www.simulations-plus.com>), etc. The need remains to validate the QSAR sub-models or the *in vitro* assays used to assign a PBTK model’s parameter values. Obviously, the quality of those inputs conditions the validity of the PBTK model which uses them. The validation of those sub-models and *in vitro* assays should be made following the relevant procedures, in the context of cosmetic ingredients. Sensitivity and uncertainty analyses can also be performed to understand which are the critical aspects of the model that might require particular attention (Bernillon and Bois 2000). Experimental or observational data are not always available to convincingly validate such complex models. ‘Virtual’ experiments simulated by varying parameters, as in sensitivity analysis, can point to important areas of future research needed to build confidence in *in silico* predictions. Formal optimal design techniques can also be used to that effect (Bois et al. 1999).

In any case, the major challenge will probably be the coupling of PBTK models to predictive toxicity models, at the cellular and at the organ level. Liver models are being developed (Yan et al. 2008; Orman et al. 2010), but their predictive power is far from established for chronic repeated dose toxicity.

Inventory of *in vivo* methods currently available

Several specific tests for studying the toxicokinetics of substances *in vivo* are described in Annex V to Directive 67/548. The OECD guideline 417 describes the procedure in more detail. The OECD guideline also states: “Flexibility taking into consideration the characteristics of the substance being investigated is needed in the design of toxicokinetic studies.” It should be mentioned that the OECD guideline 417 has been recently updated and adopted (July 2010) with the inclusion of *in vitro* and *in silico* methods. With the exception of dermal absorption, detailed data on the toxicokinetics including the metabolism of cosmetic ingredients is currently of limited importance and not requested. Such additional information is only required for cases where specific effects, seen in standard *in vivo* animal tests, have to be clarified and their relevance to humans must be proven.

For example, toxicokinetics (TK) could aid in relating concentration or dose to the observed toxicity, and to aid in understanding mechanism of toxicity. Important goals are the estimation of systemic exposure to the test substance, identification of the circulating moieties (parent substance/metabolites), the potential for accumulation of the test substance in tissues and/or organs and the potential for induction of biotransformation as a result of exposure to the test substance. Additionally, toxicokinetic studies may

provide useful information for determining dose levels for toxicity studies (linear vs. non-linear kinetics), route of administration effects, bioavailability and issues related to study design.

As described in more detail in sections “[Available non-animal methods to derive values for absorption, distribution, metabolism and excretion \(ADME\)](#)” and “[Inventory of alternative methods](#)”, there exists a number of in vitro and/or in silico methods to study many of these TK processes. For example, in vitro biotransformation models available (e.g. hepatocytes in suspension or culture) are used to provide results considered relevant for risk assessment. The same holds true for in vitro/in silico results for oral absorption where information on chemical structure (e.g. QSAR) and physical and chemical properties (e.g. logPow) may also provide an indication of the absorption characteristics. Data on in vitro protein binding may also be considered if relevant for risk assessment.

Inventory of alternative methods

Currently used in vitro guideline

To date, only one in vitro test addressing toxicokinetics is covered by an OECD test guideline. This is the guideline on in vitro dermal absorption (OECD 428, adopted on February 2004) where the principles of this method are described (OECD 2004b). The guideline is accepted by the SCCS (SCCNFP/0750/03, Final). A guidance document of the SCCS is complementing this guideline (SCCS/1358/10).

Non-validated human in vitro/in silico approaches

Test systems to measure bioavailability and in vitro biotransformation are available (as described in the specific subchapters) and routinely used for specific in-house purposes, mainly in pharmaceutical companies. For some of them, extensive sets of data are available, demonstrating their importance to produce specific qualitative and quantitative information on various pharmacokinetic characteristics. Regulatory authorities have recognised that in vitro systems are helpful in addressing especially potential biotransformation-related issues during drug development. The application of in vitro systems for biotransformation (e.g. in microsomal preparations or isolated hepatocytes) has been described in guidance documents on studies of drug–drug interactions by US (US FDA-CDER 1997, a revised draft guideline has been published in 2006) and European authorities (EMA 1997, a revision of the latter is currently under public consultation). The recently revised OECD guideline 417 (July 2010) foresees the use of in vitro and in silico methods.

By applying an exposure-based tiered approach, there would be no need to analyse the biotransformation of a cosmetic ingredient if the chemical had insignificant bioavailability or even if there is no toxicological relevance. The selection of the most appropriate in vitro models for determining absorption is therefore crucial for cosmetic ingredients. Once the absorption and the potential toxicological relevance is demonstrated, then further testing and toxicokinetic information would be necessary.

Ideally, in silico and in vitro methods should use metabolically competent human cells and/or tissues to model human toxicokinetic processes to avoid any need for species extrapolation, as recommended by the ECVAM Toxicokinetics Working Group (e.g. Coecke et al. 2006). However, the limited availability of human cells and tissues, and ethical concerns which are often raised, should be taken into account, although the use of human recombinant enzymes, transgenic cells in vitro and the possibility to cryopreserve human hepatocytes are of great help.

In this respect, it should be noted that human genetic polymorphisms of biotransformation enzymes and transporters are not covered in conventional toxicological animal approaches. The use of human cells (or subcellular fraction), recombinant enzymes and transgenic cells in vitro are the first step in trying to pick up some well-known genetic polymorphisms. This information might be useful for the risk assessor and needs to be incorporated into a tiered strategy for toxicokinetics. The issue is of importance to drug development and therapy, and for other chemicals, but with respect to cosmetics, data could be less relevant. Similarly, “barriers”, such as the BBB, BTB and BPB, have been considered to be of minor importance in the context of cosmetics, although in individual cases their role may need clarification.

Non-validated human in vivo approaches

The microdosing approach, which makes use of extremely sensitive detection techniques such as accelerator mass spectrometry and LC–MS/MS, has been employed as a first-to-man experiment to elucidate the pharmacokinetics of pharmaceuticals (Coecke et al. 2006; Hah 2009; Lappin and Garner 2005; Oosterhuis 2010; Wilding and Bell 2005). However, the need to conduct short-term animal toxicity studies before employing microdosing would block application in the IR situation. Interestingly, a possible approach could be to combine it with the TTC concept. In principle, human microdosing could possibly obtain ethical approval by keeping the total dose below the relevant threshold in TTC terms, although a clear difference in the cost/benefit ratio between pharmaceuticals and cosmetics should be taken into account. Usually, an amount somewhere in between 1 and 100 µg is administered

(<http://www.nc3rs.org.uk/downloaddoc.asp?id=339&page=193&skin=0>). If the chemical is not a genotoxic compound (sufficient *in vitro* methods available) and not an organophosphate, the lowest threshold for exposure below which adverse effects are unlikely is 90 µg/day. Acknowledging that this threshold was based on lifelong exposure, it can be argued that this might be a promising approach for further consideration. In this evaluation, the issue of exposure route should be included, because the current TTC concept is completely based on oral toxicity studies.

Imaging techniques to study both the kinetic and dynamic behaviour of pharmaceuticals or pharmaceutical-associated materials in *in vivo* conditions in humans are advancing rapidly (Agdeppa and Spilker 2009; Péry et al. 2010), but similarly as with the microdosing concept, at present it is difficult to see whether imaging techniques would become tools for cosmetics risk assessment related research.

Current developments in model systems

Organotypic culture models applicable to the intestinal, pulmonary barriers and the blood–brain barrier are being actively developed. However, these are laborious experimental systems, which are not easy to handle, and are currently restricted to mechanistic investigations or specific questions facilitating *in vitro*–*in vivo* comparisons (Garberg, et al. 2005; Prieto et al. 2010; Hallier-Vanuxeem et al. 2009).

Recent developments on microfabrication technologies coupled with cell cultures techniques have allowed for the development of “cells on a chip” (El-Ali et al. 2006; Hwan Sung et al. 2010) that have been used to mimic biological systems and even as a physical representation of a PBTK model. For example, Viravaidya et al. (2004) and Viravaidya and Shuler (2004) developed a four-chamber microscale cell culture analogue (µCCA) containing “lung”, “liver”, “fat” and “other tissue” used to study the role of the naphthalene metabolism in its toxicity and bioaccumulation using cultures of L2, HepG2/C3A and differentiated 3T3-L1 adipocytes. Tatosian and Shuler (2009) studied the combined effect of several drugs for cancer treatment using HepG2/C3A as liver cells, MEG-01 as bone marrow cells and MES-SA as uterine cancer cells and MES-SA/DX-5 as a multidrug resistant variant of uterine cancer. They showed that a certain combination could inhibit MES-SA/DX-5 cell proliferation and using a PBTK model of their device they were able to scale-up to calculate doses for *in vivo* trials. Finally, Chao et al. (2009) using a similar approach and human hepatocytes showed that they could predict the *in vivo* human hepatic clearances for six compounds.

Steps or tests with novel or improved alternative methods needed

Since only in cases where a cosmetic ingredient is bioavailable following dermal, oral or inhalation exposure, further tests on systemic and not just local toxicity will be necessary, the priority for additional efforts has to be given in providing reliable alternative methods to assess the bioavailability after oral and inhalation exposure. Several efforts have been undertaken to improve the reliability of alternative test methods available assessing absorption via gut, but still more work would be required. At present, extensive experience does not exist with *in vitro* systems suited to measure absorption through the lung alveoli. The systems used in house have some disadvantages, and the performance of a three-dimensional *in vitro* culture with pulmonary cells has not yet shown for this purpose. Hence, in the light that this route of exposure is important for cosmetics, efforts are necessary to develop systems for the purpose of measuring pulmonary absorption.

At the same time, it will be essential to develop toxicodynamic experimental design including all toxicokinetic consideration necessary to transform *in vitro* nominal concentration–effect relationship into an *in vivo* dose–effect relationship and allow the extrapolation of *in vitro*/*in silico* data to *in vivo* dose–effect relationship. If the actual applied concentration *in vitro* could be determined by using appropriate biokinetic measures, the relevance for the extrapolation could be improved.

In order to do so, more investments should be done to have access to high-throughput validated analytical methods for compound and metabolite identification, as the indispensable first step before any toxicodynamic experimental design is planned allowing a quantitative risk assessment.

Several *in vitro*/*in silico* building blocks are available, but a wide variety of standard operating procedures (SOP) is used by different industries and CROs. Alternative methods are available, but no effort up to now has been made in order to get the most reliable version of such SOP accepted by regulators.

The development of *in vitro*/*in silico* methods dealing with biliary excretion and renal excretion is felt to be essential to progress to a full 1R replacement strategy based using as integrative tool the PBPK models.

Recommendations

Given the scenario of the development of compounds/products based on non-*in vivo* animal testing strategies, toxicokinetics becomes the cornerstone in the risk assessment under 1R conditions. Toxicokinetic information has to be available upfront for assessing the need for further

testing dependent on the bioavailability, to plan *in vitro* toxicodynamic testing, and together with biokinetic *in vitro* data will allow to relate the *in vitro* information of the concentration–effect(s) relationship of the substance to an *in vivo* dose–effect relationship.

The following recommendations are given to pave the way forward in the field of toxicokinetics under the 1R scenario.

Firstly, absorption through the lung and excretion (via the kidneys and the biliary route) are the two processes within the ADME processes (absorption, distribution, metabolism, excretion), which have been identified as knowledge gaps.

Secondly, physiologically based toxicokinetic (PBTK) models are ideally suited for the integration of data produced from *in vitro* testing systems/*in silico* models into a biologically meaningful framework, and for the extrapolation to *in vivo* condition. However, even if proof of concept has been provided for the strategy how to proceed there is presently not much experience, and hence, further development and refinement is necessary. One could envisage building a publicly available user friendly tool for PBTK modelling with a repository of examples to support the use of the tool. This would promote not only PBTK as a tool for risk assessment but also the concept of risk assessment under 1R scenario as a whole. Furthermore, the working group on toxicokinetics restricted itself to kinetic aspects because it felt not charged also to explore aspects of toxicodynamics. However, it should be emphasised that the link between the results of *in vitro* effect testing, corrected on the basis of biokinetic measurements, and the PBTK modelling is by modelling the *in vitro* responses by toxicodynamic modelling (Dahl et al. 2010). Hence, the crosstalk between toxicologists measuring effects and toxicologists/scientists performing kinetic or kinetic/dynamic modelling is fundamental (see Figs. 1, 2).

Thirdly, in the areas where *in vitro*/*in silico* methods are available, it should be considered whether the conventional validation procedure is the most efficient way forward. Recognising that only one *in vitro* toxicokinetic method is accepted at the OECD level (OECD guideline 428 for *in vitro* dermal absorption), it should be considered whether and to what extent alternative methods could be utilised. It could be envisaged to work in an expert consensus procedure by collecting methods, assessing them according to test quality criteria and ranking them. Finally, by consensus a standard operating procedure (SOP) could be derived and the reliability of the method could be tested on a small sample of compounds with properties relevant for cosmetics. *In vitro* protein binding, *in vitro* metabolism and clearance, and *in vitro* oral absorption may be valid examples for this approach. Currently, ECVAM is carrying out a formal reliability check in the context of an

international validation trial of 2 metabolic competent human hepatic systems (cryopreserved human hepatocytes and the Human HepaRG cell line) for several phase 1 biotransformation CYP isoforms.

Concerning the available *in silico* methods (e.g. tissue distribution), it seems necessary to explore whether substances used in cosmetics are in the chemical space of the substances which have been used to develop and validate the algorithms. If not, adjustment or even new development of algorithms has to be undertaken.

Finally, it must be clarified that all the exercises for the process of validating and of finding acceptance needs financial resources other than research, because this activity cannot be seen as a research activity but is rather a standardising activity. There will be the necessity to involve institutions which have experience and work on standardising issues (e.g. experience from pharmaceutical companies which already use a variety of alternative methodologies).

There are some fields which have not yet been considered in depth. The most important is the field of nanoparticles. We are aware of the fact that nanoparticles are currently used in cosmetic products applied to the skin. As far as we know, the presently available data show that from the products on the market absorption to the general circulation (i.e. internal exposure) does not take place. We however know that absorption through the lung alveoli may occur. We would recommend that a special working group should be set up to deal with the issues of nanoparticles in the field of cosmetics taking into account the regulations in other fields (e.g. industrial chemicals) and what has been already considered by other institutions (e.g. OECD).

Conclusions

Under the 1R scenario which has to be envisaged to be in place from 2013 on, the risk assessment of cosmetics is faced with a radically altered situation. In the current paradigm of risk assessment, the external exposure (in mg/kg/day) is compared to the dose for an observable effect (at the no-effect-level in mg/kg/day adjusted with appropriate assessment factors). In the old paradigm, kinetic and dynamic considerations help to understand the mode of action/interspecies differences. In the new framework, knowledge on the toxicokinetic behaviour of a substance becomes the first important piece of information.

Information on toxicokinetics under the 1R is essential to address the following three major issues:

- (A) It is essential to know, whether a substance will be bioavailable by one of the relevant uptake routes: only in cases where a cosmetic ingredient is

bioavailable following dermal, oral, or inhalation exposure, further tests on systemic and not just local toxicity will be necessary.

- (B) In order to relate toxicodynamic information from non-animal-testing (1R) to real-life situation relevant for humans, it is necessary to transform the *in vitro* actual concentration–effect relationship into an *in vivo* dose–effect relationship. Physiologically based toxicokinetic modelling is the indispensable tool to enable the transformation.
- (C) In order to plan the experimental design for the *in vitro* dynamics experiments, it is essential to know whether the cells or tissues are exposed to the parent compound and/or its metabolites. *In vitro* data on metabolism support this decision.

In addition, *in vitro* biokinetic data recorded during *in vitro* toxicity testing will be crucial to derive the actual *in vitro* concentrations: indeed, nominal applied concentrations may greatly differ from the intracellular concentration due to altered bioavailability or to physiological cellular processes. In repeated treatments for prolonged time of exposure, to mimic exposure to cosmetic products, the uncertainty about the actual level of exposure of cells *in vitro* is greatly enhanced.

The development of specific and sensitive analytical methods will be the first step to obtain the necessary toxicokinetic information. Currently, there exists a number of *in vitro* and *in silico* methods to cover different aspect of the toxicokinetics processes. Although for some of them a further development/improvement is still necessary, some others of the existing methods are already well developed, but, with the notable exemption of the *in vitro* dermal absorption for which an OECD guideline exists, they are non-validated.

Regarding the validation of the non-validated testing methods, it should be considered whether the conventional validation procedure is the most efficient way forward and whether alternative methods could be utilised. It could be envisaged to work in an expert consensus procedure to set up Standard operating procedures by consensus and validation could be performed in testing the reliability of the methods with compounds possessing properties relevant for cosmetics. The appropriateness of available *in silico* methods (e.g. tissue distribution) has to be explored for substances used in cosmetics with respect to their location in the chemical space. The non-availability of methods to produce *in vitro* data on the absorption after inhalation exposure and on excretion have been identified as the major data gaps.

We underscore the importance of PBTK (PBPK) modelling as a necessary tool to organise and integrate the input from *in vitro* and *in silico* studies. In addition, we would

like to recommend including also toxicodynamic modelling in the chain from *in vitro* test results to the *in vivo* dose–effect relationship. We also recommend supporting the development of building a publicly available user friendly tool for PBPK modelling with a repository of examples.

A special working group, probably in collaboration with other concerned agencies such as European Food Safety Authority (EFSA), should be set up to deal with the issues of nanoparticles in the field of cosmetics taking into account the regulations in other fields (e.g. industrial chemicals) and what has been already considered by other institutions (e.g. OECD).

Given the best working conditions, including resources in money and manpower, it could be predicted that the improvement of the existing methods and the development of *in vitro* methods for renal excretion and absorption via the inhalation route would take 5–7 years; an integrated approach linking the results from *in vitro/in silico* methods with physiologically based toxicokinetics in order to characterise different steps involved in toxicokinetics would take a considerably longer time.

Although animal toxicokinetic models are already rarely used in the context of cosmetics and consequently the impact of the 2013 deadline will be greatly diminished, the WG emphasises that toxicokinetics is the first step in the non-animal testing strategy for cosmetics, considering a decision tree based on systemic bioavailability and on the need to integrate biokinetics into toxicity testing.

Skin sensitisation

Executive summary

Skin sensitisation is the toxicological endpoint associated with chemicals that have the intrinsic ability to cause skin allergy. To meet the challenge of the Cosmetics Regulation, we must replace the need to generate new animal test data for Cosmetics Industry risk assessments for this endpoint. Hazard characterisation data (e.g. dose–response) for an ingredient, specifically its relative potency, i.e. the chemical's power/strength to induce skin sensitisation, is used in combination with the expected human exposure to that ingredient to predict the risk to human health. For skin sensitisation risk assessment, the ability both to identify and characterise the sensitisation potential and relative potency of chemicals enables safe levels of human exposure to be predicted in the context of product use. Consequently, a non-animal approach capable of reliably predicting skin sensitiser potency is required to ensure consumer safety/continued innovation within the Cosmetics Industry. This report is a critical evaluation of the

ability of available non-animal test methods to generate information that could be used to inform skin sensitiser potency predictions (and therefore quantitative risk assessments for cosmetics). Several non-animal methods are being developed for hazard identification to support hazard classification and labelling; however, data from these test methods alone will not be sufficient for risk assessment decision-making. Therefore, we focus here upon evaluating non-animal test methods for application in risk assessment, based upon:

- mechanistic relevance to skin sensitisation
- contribution to potency determination
- evidence of reliability (i.e. robustness, reproducibility & transferability)

Table 1 displays generic non-animal test methods aligned to key mechanistic pathways of skin sensitisation [skin bioavailability (1), haptentation (2), epidermal inflammation (3), dendritic cell activation (4) and migration (5) and T-cell proliferation (6)], together with a view on when they are likely to be available for pre-validation studies. It is unlikely that full replacement of the current in vivo assays will require all methods and mechanistic steps listed in Table 1. However, at present it is not possible to predict which combinations may be required to derive potency information for individual chemicals and exposure scenarios. Consequently, a tool box covering a range of mechanistic steps has been reviewed which could provide the required information.

Please note that the estimated timelines for entry into pre-validation do not consider the time required for test method reliability to be evaluated, nor the time required to

establish the value of test data for risk assessment decision-making.

The conclusion of the expert group was that the most positive view of timing is as follows:

- 2013*: No full replacement of in vivo methods available, although hazard identification without potency information might be possible, allowing for the identification of non-sensitisers.
- 2017–2019*: Scientific ability to make skin sensitisation risk assessment (RA) decisions using a toolbox of non-animal test methods for all Cosmetics Industry ingredients & exposure scenarios.

*Note that the replacement timeline is based on the assumption that optimal conditions are met (i.e. all necessary resources are made available) and that the studies undertaken will have successful outcomes. In addition, the replacement timeline does not consider the time required for regulatory acceptance, but does consider the time required, typically 2–3 years, for demonstration that a non-animal test method is robust, relevant and delivers information useful for the determination of the intrinsic relative potency of sensitising chemicals and so is of value for risk assessment decision-making. Furthermore, although the timeline is based upon the premise that non-animal test methods that are predictive for each mechanistic step will be available, it is unlikely that information on every mechanistic step will be required to inform all risk assessment decisions. Therefore, it is expected that the scientific ability to inform skin sensitisation decisions without animal test data for *some* ingredients and exposure scenarios should be feasible ahead of 2017–2019.

Table 1 Toolbox for skin sensitisation risk assessment: estimated timelines

Mechanistic step ^a	Non-animal test methods	Estimated timeline for entry into pre-validation
1	Skin bioavailability/metabolism	2010–2017 ^c
1 + 2	In silico chemistry	N/A ^b
2	Peptide reactivity (>6 methods)	2009–2011
2 + 3	Keratinocyte cultures, including cell-based reactivity (>3 methods)	2010–2012
3 + 4	Keratinocyte/dendritic cell co-culture (>1 method)	2011–2012
3	3D reconstituted skin models (>1 method)	2010–2011
3 + 4	3D reconstituted skin models + dendritic cells (>1 methods)	2012–2015
4	Dendritic cell activation (>4 methods)	2009–2012
5	Dendritic cell migration (>1 methods)	2012–2015
4 + 6	T-cell proliferation (>2 methods)	2015–2017

The dates for entry into pre-validation refer to the earliest and latest date for the methods described

^a The numbers refer to the main steps shown in Fig. 8

^b Not applicable—in silico models should be evaluated in accordance with the OECD principles for the validation of (Q)SARs (OECD 2007a)

^c This timeline reflects the span of estimated timings for assays for the other mechanistic steps in which bioavailability is incorporated

Information requirements for the safety assessment of cosmetic ingredients and how this information is currently obtained

Introduction/description of skin sensitisation and mechanisms

Skin sensitisation is the toxicological endpoint associated with chemicals that have the intrinsic ability to cause skin allergy, termed allergic contact dermatitis (ACD) in humans. This adverse effect results from an over reaction of the adaptive immune system and thus involves two phases, the induction of sensitisation which is followed upon further contact with the sensitising chemical by the elicitation of allergy symptoms. The elicited symptoms of ACD are usually largely confined to the area where the allergen actually touches the skin and include the following: red rash (the usual reaction) and blisters, itching and burning skin. Detailed reviews of the mechanistic aspects of skin sensitisation/allergic contact dermatitis can be

found elsewhere (Rietschel and Fowler 2008; Vocanson et al. 2009; Basketter and Kimber 2010a). The main steps are shown in Fig. 8 and involve

1. Skin bioavailability—the extent to which the compound reaches the site for haptentation.
2. Haptentation—the covalent binding of the chemical sensitiser to skin protein.
3. Epidermal inflammation—the release of pro-inflammatory signals by epidermal keratinocytes.
4. Dendritic cell (DC) activation—the activation and maturation of skin-associated DCs (i.e. Langerhans cells) in response to the combined effects of steps 1 and 2, including maturational changes to the DCs.
5. DC migration—the movement of hapten–peptide complex bearing dendritic cells from skin to the draining lymph node.
6. T-cell proliferation—the clonal expansion of T cells specific for the hapten–peptide complex.

It is important to recognise that contact allergens vary widely in their relative skin sensitising potency, i.e. in their intrinsic capacity for inducing skin sensitisation, some being very strong, whereas others are much less potent. The relative sensitising potency is defined as the ease with which a chemical is able to induce sensitisation compared to benchmark allergens for which information on potency in humans is already available. The more potent the chemical allergen, the lower the dose required to cause the acquisition of skin sensitisation. As described in more detail in the following sections, an accurate evaluation of the relative potency of skin sensitising chemicals is the key requirement to fully inform the risk assessment process. Notably, for the purpose of hazard identification, sensitising potency information has to be taken into account according to the recently adopted Globally Harmonised System of Classification and Labelling of Chemicals (GHS; Anon 2003).

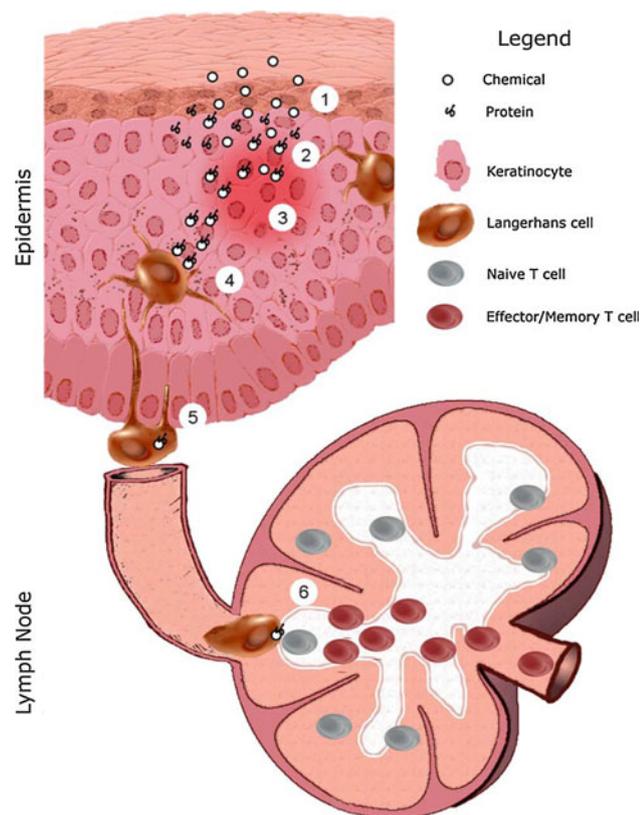


Fig. 8 Main steps in the mechanism of skin sensitisation induction. The numbers correspond to the steps described in the text. (1) Skin bioavailability, (2) haptentation, (3) epidermal inflammation, (4) DC activation, (5) DC migration, (6) T-cell proliferation. This figure contains elements of an image in the public domain from the National Cancer Institute

Inventory of animal test methods currently available

Predictive testing to identify and characterise substances causing allergic contact dermatitis historically has been based on animal tests. The standard and accepted skin sensitisation test methods, for which OECD guidelines are available, include the guinea pig maximisation test (GPMT) according to Magnusson and Kligman, the occluded patch test of Buehler and the mouse local lymph node assay (LLNA). The guinea pig regulatory test protocols have not been designed to deliver potency information, which is in contrast to the LLNA.

Guinea pig maximisation test (GPMT) In the GPMT, guinea pigs are successively exposed to the test substance

by intradermal injection (with and without Freund's complete adjuvant as immune enhancer) and topical application by occlusion (induction exposure). Following a rest period of 10–14 days (induction period), the animals are exposed dermally to a challenge dose using 24-h occlusion (Magnusson and Kligman 1970). The extent and degree of skin reactions to this challenge exposure is then compared with control animals. A rechallenge treatment may be considered 1–2 weeks after the first challenge to clarify equivocal results. Test substances are regarded as skin sensitisers when at least 30% of the animals show a positive response. Details of the method are given in the respective guidelines (OECD 1992; EU 2008).

Buehler guinea pig test In the Buehler test, guinea pigs are repeatedly exposed to the test substance by topical application under occlusion (induction exposures). Following a rest period of 12 days (induction period), a dermal challenge treatment is performed under occlusive conditions. The skin reactions to the challenge exposure are compared with the reaction in control animals (Buehler 1965). A rechallenge treatment may be considered 1–2 weeks after to clarify equivocal results. Test substances are regarded to be skin sensitisers when at least 15% of the animals show a positive response. Details of the method are given in the respective guidelines (OECD 1992; EU 2008).

Mouse local lymph node assay (LLNA) In the LLNA, the test substance is applied to the dorsum of the ears of mice for 3 consecutive days. On day 5, tritiated thymidine is injected intravenously as a radioactive label for the measurement of cell proliferation. Five hours later, the auricular lymph nodes are excised and the incorporated radioactivity counted (Kimber and Basketter 1992). An increase in lymph node cell proliferation (stimulation index, SI) compared to concurrent vehicle-treated control animals indicates sensitisation (thus, the LLNA focuses on the induction of sensitisation, not elicitation). The test is positive when the SI is ≥ 3 . The estimated concentration of a substance necessary to give a threefold increase is the EC3 value (Basketter et al. 1999). Details of the methodology are given in the relevant guideline (OECD 2002; EU 2008). A reduced version of the LLNA (rLLNA) compared to the standard protocol has been proposed (Kimber et al. 2006). The rLLNA has been recently accepted by the OECD in an update of TG 429. Two new test guidelines for the non-radioactive modifications of the LLNA, the LLNA: DA method (OECD TG 442A; OECD 2010b) and the LLNA: BrdU-ELISA assay (OECD TG 442B; OECD 2010c) have been adopted by the OECD in 2010.

Information supplied by these tests and its use for risk assessment

Risk assessment for skin sensitisation relies on the same elements that are used in general toxicology practice: hazard identification, dose–response assessment, exposure assessment and risk characterisation. For decades, the standard guinea pig tests (Buehler test, GPMT), developed for the identification of skin sensitisers have been used as reliable predictors of allergic hazard (e.g. Andersen and Maibach 1985). However, these assays do not provide information on dose responses or thresholds for sensitisation, because only a single test concentration is used for the induction and challenge treatment (Basketter et al. 2005a; van Loveren et al. 2008). In principle, it is possible to obtain data on skin sensitisation potency from guinea pig tests by using a multidose regime for induction and/or challenge treatment, but this has rarely been done (Andersen et al. 1995; Frankild et al. 2000). In practice, the information from guinea pig tests has been interpreted for risk assessment purposes by considering these aspects:

- Physicochemical data (e.g. the octanol–water partition coefficient K_{ow}) of the test compound to get an estimate of its bioavailability.
- Test concentrations used for induction and challenge treatment(s) with emphasis on the exposure route (topical intradermal).
- Examination of whether a response may be a false positive, e.g. use of rechallenge.
- Estimation of potency based on induction and elicitation concentrations, the frequency and intensity of positive reactions and comparison to well-characterised benchmark allergen(s).
- Comparison of the estimated potency with general use conditions (concentration, frequency, duration).
- Applying a safety factor if necessary.

In contrast to the above, the LLNA has been recognised as more capable in terms of the predictive identification of the relative potency of skin sensitising chemicals (Basketter et al. 2005a; van Loveren et al. 2008). In particular, the use of dose–response data to estimate the concentration necessary to produce a threefold increase in stimulation (the EC3 value) has been deployed as a marker of relative potency (Griem et al. 2003; Schneider and Akkan 2004; Basketter et al. 2005c, 2007a; Basketter 2010).

Current risk assessment

Cosmetic products have multiple uses by many millions of consumers on a daily basis. To determine whether a skin sensitising chemical can be used safely in any particular product, its potency must be considered in the context of its

concentration in the product and the predicted human exposure scenario (Gerberick et al. 2001; Felter et al. 2002; Felter et al. 2003; Api et al. 2008). Typically, the predicted potency of the ingredient in the LLNA is deployed in a quantitative risk assessment (QRA) approach reviewed elsewhere (Felter et al. 2002; Felter et al. 2003), so only a brief description will be given here. Several studies have demonstrated a correlation between sensitisation induction thresholds in the LLNA (the EC₃ value) with the relative skin sensitising potency of chemicals in human predictive testing, notably the human repeated insult patch test (HRIPT) (Griem et al. 2003; Schneider and Akkan 2004; Basketter et al. 2005c, 2007a, 2008a; Basketter 2010). These analyses covered a broad spectrum of sensitisation mechanisms and potencies. Thus, a no expected sensitisation induction level (NESIL) in the HRIPT can be predicted from the mouse LLNA. Then, three different categories of uncertainty factors (UF) are applied to the NESIL values to capture the potential for inter-individual differences, vehicle or product formulation differences, and exposure considerations. A default value is applied for each UF (Felter et al. 2002; Felter et al. 2003; Api et al. 2008). This modification of the NESIL using UFs provides a theoretical safe level of exposure to material which can then be compared with the predicted consumer exposure to the material in the product. The outcome of the QRA is also interpreted alongside any existing information, such as history of use or clinical data relating to the material in question (or similar benchmark materials), in order to reach an overall risk-based safety decision. Practical examples of this approach have been published (e.g. Api et al. 2008; Basketter et al. 2008b; Basketter 2010).

It is important at this point to note that predictive skin sensitisation testing in humans, e.g. the human repeated insult patch test (HRIPT), should not be part of methods for cosmetic testing and is generally discouraged (SCCNFP 2000; SCCP 2008; van Loveren et al. 2008; Basketter 2009).

The threshold of toxicological concern (TTC) concept is a pragmatic risk assessment tool that establishes a human exposure threshold for chemicals without specific toxicity data (Kroes and Kozianowski 2002). This has been applied for decades in cosmetic product safety evaluation and recently has been crystallised in two publications concerning its application to skin sensitisation (Safford 2008; Keller et al. 2009).

Non-animal tools to inform the cosmetic industry for risk assessment

To determine whether a chemical has the potential to induce skin sensitisation, non-animal test methods are being developed which reflect the key mechanisms

involved in skin sensitisation. It has been proposed that no single approach could predict sensitiser potency, and that the integration of multiple forms of non-animal data/information would be necessary (Jowsey et al. 2006; Basketter and Kimber 2009; Roberts and Patlewicz 2010). Consequently, the integration of some or all of the following categories of non-animal information will yield a new measure of skin sensitiser potency to inform the new risk assessment process: bioavailability, haptenation, epidermal inflammation, dendritic cell activation, dendritic cell migration and T-cell proliferation. In line with this, the section that follows reviews promising ongoing research, method development and method evaluation activities aligned to each of the key mechanistic steps shown in Fig. 8. Particular focus has been given to material published since the last review (Basketter et al. 2005b).

There are ongoing research projects to develop generic strategies for integrating the results obtained from these different categories. The development of integrated testing strategies (ITS) for human health endpoints, including skin sensitisation, is one focus of the OSIRIS project included under the Sixth Framework Programme funded by the European Commission. In addition, an important objective of the OSIRIS project is to develop a generic strategy for ITS including quantitative estimates of certainty. A review of ITS for skin sensitisation is provided in the JRC Scientific and Technical report by Patlewicz and Worth (2008), prepared as a contribution to the OSIRIS project. Several literature-reported ITSs for skin sensitisation are reviewed here as for example the ITS for skin sensitisation customised by Grindon et al. (2006) highlighting that before conducting new *in vivo* test, *in vitro* and *in silico* data shall be considered. A non-animal strategy exploiting QSAR approaches which involve applying mechanistic principles is also presented (Roberts and Aptula 2008). The authors describe the determinants of skin sensitisation potential from a chemical perspective arguing that the rate-limiting step involves the covalent binding of a chemical electrophile to a protein nucleophile. Such electrophiles can be classified into a limited number of reaction mechanistic domains within which QMM (Quantitative Mechanistic Model) may be derived using the RAI (Relative Alkylation Index) approach which relates reactivity and hydrophobicity to sensitisation potential.

Bioavailability Before moving on to discussions of non-animal approaches, it is useful to be reminded that the *in vivo* models that have an intact skin barrier are believed to model differences in how different chemicals penetrate into and are metabolised by the skin in a manner that reflects what occurs in humans. Even though this assumption has not been extensively explored, there are ongoing activities to predict skin bioavailability using non-animal test

methods and to establish the metabolic competency of human skin equivalent models as a potential model for in vivo skin metabolism (Gibbs et al. 2007; Pendlington et al. 2008; Kasting et al. 2010). However, it is reasonable to anticipate that the use of in vitro methods will be associated with a markedly lower, or even absent, impact of bioavailability and metabolic considerations on the prediction. A recent review examined this topic in detail, without identifying how the problem could be resolved (Basketter et al. 2007c). In a pragmatic sense, and as will be mentioned below, quantitative structure activity relationships (QSARs) often include logP (the octanol/water partition coefficient) as a variable, where it is considered to reflect the epidermal disposition of a chemical (Roberts and Williams. 1982). This variable was also included in a first effort to show how in vitro skin sensitisation data might be combined in practice (Natsch et al. 2009). Whether logP proves to be appropriate as an indicator of epidermal availability for future in vitro methods, particularly with respect to potency prediction, remains unknown.

Ultimately, it may be the nature of the in vitro assay(s) which prove successful in informing the identification and characterisation of skin sensitising chemicals that will themselves determine the type of bioavailability information necessary to complement them (Davies et al. 2010). Furthermore, to understand the sensitising potency of a chemical in the absence of in vivo data, it will be important to establish whether metabolic activation/inactivation could occur upon skin exposure. (Karlberg et al. 1999; Smith and Hotchkiss 2001; Ott et al. 2009).

A final word on metabolic activation/inactivation of skin sensitising chemicals: there is virtually no information on what metabolic capabilities are important in human contact allergy. This means that, at present, it is not possible to determine what metabolic capacity would be appropriate in any predictive model. Consequently, there is no value in listing the presence of absence of such capabilities in current in vivo or in vitro models. In practice, the assays will have to be evaluated by the practical experience of sensitising chemicals thought to need metabolic intervention in order to function.

Mechanistic chemistry and in chemico reactivity (mechanistic step 2) Theoretical chemistry: Research dating back more than seven decades has established a strong mechanistic understanding of skin sensitisation; however, some knowledge gaps still exist, including a clear understanding of the nature and location of carrier proteins (Divkovic et al. 2005; Roberts et al. 2007). In brief, the ability of a chemical (either as a direct acting or after autoxidation/metabolism) to react covalently with a ‘carrier protein’ is a major determinant factor in its ability to act as a skin sensitizer (Aptula and Roberts 2006; Roberts and

Aptula 2008). The ability of the causative chemical species to react covalently with the carrier protein is related to either electrophilic reactivity or both electrophilic reactivity and hydrophobicity. A chemical with skin sensitisation potential is either directly reactive (electrophilic) with protein nucleophiles or it requires activation (either metabolic or autooxidative) to make it into a reactive electrophile (i.e. it is a proelectrophile) (Aptula et al. 2007). To predict and characterise the sensitisation potential of a chemical (i.e. make a yes/no prediction and provide some quantitative measure of relative potency), the following chemistry-based information is required: (a) mechanism of action (the chemical nature of its electrophilic reactions, i.e. assignment to its reaction mechanistic domain) (Divkovic et al. 2005; Aptula and Roberts 2006; Roberts et al. 2007), (b) hydrophobicity (usually expressed as logP) models bioavailability at the location where the protein-binding reaction leading to sensitisation occurs. For some reaction mechanistic domains logP is not required, bioavailability being simply related to the dose/unit area (which in turn is directly related to concentration in the LLNA), but in others, where bioavailability depends on partitioning into a lipid phase where reaction takes place, logP is required in combination with reactivity, (c) reactivity (ideally kinetic, models the protein binding of the bioavailable sensitizer).

In silico tools for estimating skin sensitisation: Several expert systems for predicting skin sensitisation are freely available, to mention some:

1. The Toxtree software allows to classify the chemical into its mechanistic domain (Aptula and Roberts 2006; Enoch et al. 2008). This grouping will allow a user to group chemicals into a mechanistic category. The data for the chemicals within the category can then be used in trend analysis/read across to fill any skin sensitisation data gaps within the category. Toxtree, along with a database of peer-reviewed QSARs and information on the validation status (according to the OECD principles for the validation of QSARs; OECD 2007a) of these models and how to report their use, are available from EU website <http://www.ecb.jrc.ec.europa.eu/qsar/>.
2. The OECD (Q)SAR Application Toolbox allows users to apply structure–activity methods to group chemicals into categories and then to fill data gaps by read-across, trend analysis or external (Q)SARs. The key step in using the Toolbox is the formation of a chemical category (OECD 2007b), which is a group of chemicals whose physicochemical and human health and/or environmental toxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern as a result of structural

similarity. Since variations in chemical structure can affect both toxicokinetics (uptake and bioavailability) and toxicodynamics (e.g. interactions with receptors and enzymes), definitions of which chemicals should be included in a category and conversely which chemicals should be excluded as well as a definition of the chemical space of the category are essential (Aptula and Roberts 2006). Filling the data gaps with the Toolbox is only possible when there is measured data for one or more chemicals within the category. Experimental results for skin sensitisation which are searchable by structure within the Toolbox include the database developed and donated to the Toolbox by the Laboratory of Mathematical Chemistry, Bulgaria. The Toolbox itself is freely available for download from the following public internet site: <http://www.oecd.org/env/existingchemicals/qsar>.

3. Statistically based models for predicting skin sensitisation have been developed within the EU-funded CAESAR project (<http://www.caesar-project.eu>), implemented into open-source software and made available for online use. These models have been developed and tested under stringent quality criteria to fulfil the principles laid down by the OECD and the final models, accessible from CAESAR website, offer a method of assessing skin sensitisation for regulatory use (Chaudhry et al. 2010). A new version of the CAESAR models (CAESAR v. 2.0) will be made freely available soon and will include some new features to obtain more reliable predictions and will allow to assess the applicability domain (AD), through quantitative and visual ways.
4. The ability to make predictions on the skin sensitisation potential based on in chemico reactivity has been recently evaluated under the UK Defra LINK project (funded by the United Kingdom Department for Environment, Food and Rural Affairs). QSARs, Quality assured databases together with integrated testing strategies decision tools for skin sensitisation, have been developed and are freely available under <http://www.inchemicotox.org/>.

Several expert systems which claim to predict skin sensitisation are commercially available, to mention some:

1. DEREK for Windows (<https://www.lhasalimited.org/derek/>) has an extensive rule base able to identify skin sensitisers. The rule base within Derek for Windows is mechanistically based taking the premise that haptation is the key event that leads to skin sensitisation. Such systems are of great benefit in supporting other predictions (for example, predictions made by read across or statistical QSARs).

2. TIMES-SS (<http://www.oasis-lmc.org/?section=software>)—Times MEtabolism Simulator platform used for predicting skin sensitisation is a hybrid expert system that was developed at Bourgas University using funding and data from a consortium comprised of industry and regulators. TIMES-SS encodes structure–toxicity and structure–skin metabolism relationships through a number of transformations, some of which are underpinned by mechanistic three-dimensional quantitative structure–activity relationships.

It must be kept in mind that none of the above approaches represents a complete replacement for the current in vivo methods nor have they undergone any formal validation (see for example Patlewicz et al. 2007; Roberts and Patlewicz 2008).

Reactivity assays: The general concept that the rate determining step in the skin sensitisation process is the reaction of the sensitiser with skin nucleophiles has led to initiatives to develop methods and to generate data on the reactivity of chemicals towards model nucleophiles representing peptide and protein nucleophiles in the skin. Empirical measures of the reactivity of chemicals with model nucleophiles such as thiol can be used to simulate the relative rates at which a reactive chemical is likely to bind to nucleophiles in cellular targets (Gerberick et al. 2008). To emphasise the nature of the reaction, it has been termed in chemico reactivity (based only on organic chemistry). With this approach, the toxicity (e.g. skin sensitisation) of a new chemical can be estimated from measured chemical data (Schultz et al. 2009). Specific protocols to achieve this aim have been presented by a few groups. The most basic method is the direct peptide reactivity assay (DPRA) which is currently undergoing pre-validation at ECVAM in respect to establishing its reliability (within and between laboratory reproducibility and transferability) and a preliminary assessment of its predictive capacity for hazard identification (Gerberick et al. 2004; Gerberick et al. 2007). A next generation DPRA is also under development, which uses horseradish peroxidase and hydrogen peroxide (HRP/P) to capture some aspects of metabolism/autoxidation and thereby incorporate chemicals that require activation (Gerberick et al. 2009). To correct for some of the limitations of the assays mentioned above, Natsch and Gfeller (2008) introduced a variant. It is based on the alternative heptapeptide Cor1-C420, and uses quantitative LC–MS to measure peptide depletion and adduct formation. Schultz and colleagues developed a method determining thiol reactivity as a system for in chemico modelling for a variety of toxicological endpoints, including skin sensitisation (Schultz et al. 2005). Interestingly, this method can be adapted to serve as a full

kinetic assay (Roberts et al. 2010; Böhme et al. 2009). Similar approaches include the extensive reactivity/adduct profiling suggested by Aleksic and colleagues (Aleksic et al. 2009). Alternatively Chipinda et al. published a simple and rapid kinetic spectrophotometric in chemico assay involving the reactivity of electrophilic sensitizers to nitrobenzenethiol (Chipinda et al. 2010).

Among other general limitations (e.g. solubility issues), metabolism and oxidation are not included in any of the above approaches, except for the next generation DPRA assay, using HRP/P, which is a promising system to capture some aspects of metabolism/autoxidation (Gerberick et al. 2009). Furthermore, most of the above test methods are based on the thiol nucleophile and so will not be good at capturing reactivity in some domains such as Schiff Bases, acyl transfer agents, where amine nucleophiles are more relevant. The DPRA and the approach suggested by Aleksic are exceptions as they also include a lysine peptide.

Epidermal inflammatory responses (mechanistic step 3) Keratinocytes (KCs) are the predominant cells in the epidermis and represent the first line of defence of the skin to xenobiotics. KCs normally lack antigen-presenting capacity; however, upon stimulation they are able to secrete a wide range of pro-inflammatory mediators and growth factors, some of which play an important role in the initiation of the immune response (Griffiths et al. 2005). In addition, KCs are metabolically competent cells and so may be responsible for the conversion of pro-haptens into reactive metabolites (Smith and Hotchkiss 2001).

It has been theorised that the irritant properties of a substance may exert an important influence on the extent to which contact sensitisation is induced (McFadden and Basketter 2000; Basketter et al. 2007b). This implies that the potency of chemicals in inducing skin sensitisation might depend on the level of cytokine production by KCs. A dose-dependent release of IL-1 α and IL-18 has been shown following exposure of the murine keratinocyte cell line HEL-30 to sensitizers (Van Och et al. 2005). Furthermore, the authors observed that the ranking of potency was similar to the ranking established using the LLNA.

More recently a concentration-dependent increase in intracellular IL-18 at non-cytotoxic concentrations of chemical was observed in the human keratinocyte cell line NCTC2455 following 24-h treatment (Corsini et al. 2009). Notably no changes in the baseline level of IL-18 were observed following treatment with respiratory allergens or irritants indicating that cell-associated IL-18 may provide an in vitro tool for identification and discrimination of contact versus respiratory allergens and/or irritants. The approach developed in the framework of the EU FP6 Sens-it-iv project is currently undergoing inter-laboratory

evaluation to assess its transferability and reproducibility in discriminating between sensitizing and non-sensitizing chemicals, as well as its ability to estimate potency (Sens-it-iv 2010).

Recent studies have suggested that the oxidative stress response pathway is a major pathway that is induced by contact sensitizers. In this pathway, the Kelch-like ECH-associated protein 1 (Keap1), which contains cysteine residues that are highly reactive to haptens, plays an important role. Natsch (2010) recently reviewed evidence for the major role of the oxidative stress response pathway after contact sensitizer exposure. From this, a test method (called KeratinoSens) to identify sensitizers was developed using an ARE reporter construct (eight ARE sequences upstream of a luciferase gene) in the HaCaT keratinocyte cell line (Natsch and Emter 2008; Emter et al. 2010). An interlaboratory trial of this test method is underway and has been submitted to ECVAM (A. Natsch, personal communication).

The SenCeeToxTM approach measures changes in gene transcription of up to 6 different genes (under the control of the ARE promoter—see above) following treatment with test chemical of either immortalised human keratinocytes (HaCaT) cells or in a proprietary reconstructed human epidermis (RHE) model. These changes in gene transcription are integrated with data from a glutathione binding test using a predictive algorithm to generate predictions to identify skin sensitizers and to perform an initial characterisation of their potency (McKim et al. 2010).

Human skin equivalents (reconstituted tissue models) (mechanistic step 3) There are two main categories of reconstituted human skin models: skin equivalents and reconstituted human epidermis (RHE). Skin equivalents have both epidermal cells (keratinocytes) and dermal substitutes (collagen and fibroblasts) in their three-dimensional (3D) structure. RHE are the most commonly used skin models and only contain the keratinocytes in a multilayered culture to mimic the outer layer (epidermis) of the skin. At present, several commercial reconstituted RHE models are available (Vandebriel and van Loveren 2010). They lack the presence of dendritic cells/T lymphocytes critical to the induction and elicitation of skin sensitisation in vivo (Vandebriel and van Loveren 2010). Recently, the Japanese researchers Uchino et al. (Uchino et al. 2009) have developed a reconstituted 3D human skin model which utilises a scaffold of collagen/vitrigel membrane and contains DCs, keratinocytes and fibroblasts (VG-KDF-Skin). The effects of non-sensitizers and sensitizers on CD86 expression and cytokine release differed from each other, which suggest that the VG-KDF-Skin model could be relevant for further development in the detection of skin sensitizers in vitro. However, as yet there have been no

method evaluation studies exploring the detection and characterisation of skin sensitisers using reconstituted human epidermal models containing immune cells.

It is possible to incorporate the commercial RHE models into an integrated testing strategy being developed to distinguish contact sensitisers from non-sensitising (Jowsey et al. 2006; Aeby et al. 2007; Basketter and Kimber 2009; Natsch et al. 2009; Sens-it-iv). It has been proposed that most contact sensitisers are also irritants and have to penetrate the stratum corneum in order to induce a general irritant alarm signal to initiate an innate immune response, which in turn triggers an adaptive immune response. As mentioned earlier, one of the factors determining whether a sensitiser is strong or weak may be related to its degree of irritancy (McFadden and Basketter 2000; Basketter et al. 2007b; Spiekstra et al. 2009).

In conclusion, the developed methodologies where the DC cells are incorporated in the reconstituted skin model are promising, but there is no ongoing method evaluation on them available to date. The use of RHE models is also promising, especially when these models are incorporated in an integrated testing strategy. In the framework of the Sens-it-iv project, a tiered testing strategy to detect skin sensitisers encompassing the use of a RHE model to evaluate the potency of chemicals has been proposed. The strategy is currently being optimised and evaluated in a ring trial involving different laboratories.

Dendritic cell responses (mechanistic step 4) Langerhans cells (LCs) are specialised immature dendritic cells (DCs) residing in the skin. These cells are able to internalise the hapten–protein conjugate and to present it to responsive T lymphocytes. If the T lymphocyte can specifically recognise the hapten conjugate and is sufficiently activated by the DCs co-stimulatory signals, then it will be stimulated to divide resulting in an expanded population of allergen-specific T lymphocytes. The key role of LCs in the skin immune response has led many investigators to exploit their use in *in vitro* tests to detect contact sensitisers. However, because of the rarity of LCs in the skin and their spontaneous maturation during the extraction procedures, other sources of antigen-presenting cells have been considered. Primary DCs derived from peripheral blood mononuclear cells (PBMCs) or from CD34⁺ progenitor cells cultured from cord blood or bone-marrow samples represent an option (Casati et al. 2005). Activation of DCs is determined by measuring changes in the expression of cell-surface markers such as CD1a, CD40, CD54, CD83, CD86, CCR7, E-cadherin and human leucocyte antigen (HLA)-DR or the release of cytokines such as IL-1 β , IL-1 α , IL-6, IL-8 and TNF- α . The use of human myeloid cell lines such as THP-1, U-937, MUTZ-3 or KG-1 overcome the donor-to-donor variability and the difficulties in

standardising the protocols for isolating and culturing human primary DCs. The different types of DC-like cell lines and the biomarkers investigated to discriminate between sensitising and non-sensitising chemicals have been the subject of a recent review (Galvão dos Santos et al. 2009).

Among the test methods based on the use of DC-like cell lines, the myeloid U937 skin sensitisation test (MUSST) and the human cell line activation test (h-CLAT) are the most advanced in terms of standardisation and number of chemicals tested. The MUSST protocol is an improved version of the one published (Python et al. 2007). Using flow cytometry, these tests monitor the induction of cell-surface markers following exposure to the chemical. In the MUSST, changes in CD86 expression in the U937 cell line are detected; in the h-CLAT modulation of both CD86 and CD54 expression are recorded (Ashikaga et al. 2006; Sakaguchi et al. 2006; Sakaguchi et al. 2009). In both assays, concentration selection is crucial for obtaining reliable results. Both methods are being evaluated together with the DPRA in an ECVAM coordinated prevalidation study in which the tests' transferability and reproducibility will be assessed in view of their future use in an integrated approach to achieve full replacement of the animal tests. However, it is vital to recognise that these methods are only under evaluation in relation to hazard identification and not for potency assessment and thus would not be sufficient in isolation to replace the need for animal test data for cosmetic industry safety risk assessment.

The analysis of gene expression changes associated with DC exposure to contact allergens has been and continues to be the subject of a number of research activities. The VITASENS[®] test method is the result of a series of transcriptomics analyses that identified 13 genes as predictive biomarkers following treatment of human DC derived from CD34⁺ progenitor cells with sensitising and non-sensitising chemicals. An initial study with 21 chemicals measured changes in gene expression using real-time PCR (Hooyberghs et al. 2008). Subsequent analysis of the set of genes revealed that the expression profiles of cAMP-responsive element modulator (CREM) and monocyte chemotactic protein-1 receptor (CCR2) displayed highest discriminating potential between sensitisers and non-sensitising after 6-h exposure. More recently 15 skin sensitising chemicals have been used to demonstrate the linear correlation between the two genes expression changes and the concentration of the compound that causes 20% cell death (IC₂₀) with LLNA EC₃ values (Lambrechts et al. 2010).

Recent investigations also suggest that cell-surface thiols may play a role in the activation of DCs by haptens (Hirota et al. 2009). Suzuki et al. reported that hapten treatment caused alterations in the levels of cell-surface thiols on the human monocyte cell line THP-1 (Suzuki

et al. 2009). Using flow cytometry techniques to detect and quantify the cell-surface thiols, they evaluated the effects of 36 allergens and 16 non-allergens. Optimal correlation with *in vivo* data was seen when the criterion for classification was set as a greater than 15% change (either a decrease or an increase) in the levels of cell-surface thiols induced by 2 h treatment of the cells with the chemical.

Keratinocyte/DC co-culture systems (mechanistic step 4)

DC activation is dependent upon KC responses following pathogenic infection or physical/chemical insult. Consistent with the pivotal role of KCs in modulating the extent of DC activation, KC/DC co-culture models have been developed for application in skin sensitisation hazard characterisation. Although several co-culture systems are currently in development (e.g. within the Sens-it-iv FP6 project) a robust, reproducible *in vitro* test method capable of accurately predicting changes in these inflammatory pathways has yet to become available.

A KC and DC co-culture model for the prediction of skin sensitisation potential has recently been described (Schreiner et al. 2007). The co-culture system is called loose-fit co-culture-based sensitisation assay (LCSA) and is composed of human non-differentiating KCs and human monocytes that differentiate to a kind of DC with the help of exogenous cytokines. The readout of the LCSA is the modulation of the CD86 surface markers following 48-h treatment with increasing doses of test substance. Cell viability is concurrently assessed by 7-amino-actinomycin D (7-AAD). Estimation of the concentration required causing a half-maximal increase in CD86 expression allows to categorise chemicals on the basis of their sensitising potential and the estimation of the concentration required to reduce viability by 50% is used to quantify the irritation potential of a chemicals. Preliminary results obtained with a limited panel of chemicals show that the test is able to discriminate between sensitising and non-sensitising chemicals with SLS and nickel correctly classified. In addition, a good correlation between the LCSA potency categories and the LLNA potency classification has been reported (Wanner et al. 2010). Further investigation is needed with a larger panel of chemicals to confirm these results.

Dendritic cell migration (mechanistic step 5) As depicted in Fig. 8, one of the key events involved in the induction of skin sensitisation is the migration of LC, from the epidermis to regional lymph nodes where they present the antigen to responsive T lymphocytes. The process of mobilisation from the skin and migration of LC is regulated by a number of cytokines and chemokines and by their interaction with LC receptors. The induced maturation of LC is characterised by the downregulation of skin-homing chemokines

receptors such as CCR1 and CCR5 and concomitant upregulation of receptors favouring the homing to the regional lymph nodes such as CCR7 and CXCL12 (Kimber et al. 2000; Ouwehand et al. 2008).

Recently an *in vitro* DC-based migration assay has been described (Ouwehand et al. 2010) based on differential MUTZ-3 derived DC migration towards either CXCL12 or CCL5 fibroblast-derived chemokines. Preliminary results show that sensitisers and non-sensitisers can be distinguished based on predominant migration of MUTZ-3 cells to either CXCL12 or CCL5.

T-cell responses (mechanistic step 6)

Naïve T-cell proliferation in response to chemical treatment is a robust indicator that a substance is immunogenic, and several publications are concerned with demonstrating the experimental feasibility of inducing a naïve T-cell proliferation *in vitro* following co-culture with chemical sensitiser-treated DCs (Hauser and Katz 1990; Guironnet et al. 2000; Rougier et al. 2000). However, the sensitivity of this approach has not been demonstrated to date, as significant and reproducible proliferative responses have generally only been detectable following stimulation with sensitisers of strong or extreme potency. In addition, the complexity of DC/T-cell co-culture protocols has historically made standardisation of these approaches labour intensive and difficult to achieve. However, new insights into the role of regulatory T cells in the modulation of sensitiser-induced T-cell proliferation are now being applied to generate next generation DC/T-cell co-culture models. For example, it was found that depletion of CD25⁺ regulatory T cells prior to measuring T-cell proliferation and IFN- γ secretion following exposure of human autologous, peripheral blood-derived DC/T-cell co-cultures treated with well-characterised sensitisers increased the probability that T-cell proliferation would be detected for sensitisers but not non-sensitisers (Vocanson et al. 2008). These results, although encouraging, are preliminary, and consequently it is widely believed that a robust and reproducible *in vitro* T-cell proliferation model will require further long-term research in method development activities.

Recent work has identified an alternative opportunity to examine, and perhaps derive a predictive method from, sensitiser-induced responses in T cells (Aliahmadi et al. 2009). Based upon this research, a method has been proposed, the CAATC (contact allergen-activated T cell) assay. The approach uses DCs from skin and characterises the sensitising potency of chemicals via DC-induced expression of lineage specific T-cell transcription factors and cytokines, including T-bet, RORC2, IFN γ , IL-17 and IL-22 (Martin et al. 2010). However, this assay is only at the earliest stage of development, so further evidence of its

utility will be necessary before any firm conclusions can be made.

Identified areas with no alternative methods available and related scientific/technical difficulties

A complete skin sensitisation risk assessment for cosmetic safety evaluation requires not only hazard identification but also potency determination for identified sensitising ingredients. Thus, *in vitro* alternatives which satisfy basic regulatory toxicology needs regarding hazard identification will not suffice to fully replace animal tests such as the LLNA. Even these *in vitro* methods for hazard identification are currently unavailable, although they are under very active development/prevalidation. Furthermore, the understanding of what determines the potency of a skin sensitiser is still incomplete, although it is recognised that reactivity is likely to be an important component. However, translation of non-animal data into information which permits a complete risk assessment decision-making remains at an early stage of development.

In addition, it is important to keep in mind that the current *in vivo* tests, while valuable tools, are themselves not perfect (e.g. Kreiling et al. 2008; Basketter and Kimber 2010b). Thus, it is not reasonable to expect replacement methods to be perfect, nor to correlate precisely with the existing *in vivo* tests.

Summary of alternative methods currently available and foreseeable time to achieve full replacement of the animal test(s)

See Table 2.

Conclusions/summary

Skin sensitisation risk assessment decisions for the safety evaluation of cosmetic products require not only hazard identification but also sensitiser potency information to allow a safe level of human exposure to be predicted for chemical sensitisers. Consequently, non-animal test methods designed to characterise sensitiser potency are required for complete skin sensitisation risk assessment decisions to be made in the absence of new animal test data. It is recognised that data from non-animal test methods developed to identify skin sensitisation hazard potential could be applied to risk assessment decision-making. For example, if the absence of a skin sensitisation hazard was identified, no further information on potency would need to be generated. However it is important to emphasise that, in isolation, non-animal tests for hazard identification will not be sufficient to replace fully the need for animal testing for this endpoint, although they might reduce the overall need.

Despite significant progress since the publication of the previous review of this field (Basketter et al. 2005b) in the standardisation of non-animal test methods for hazard identification (e.g. three test methods are currently undergoing ECVAM prevalidation), currently our ability to predict sensitiser potency without animal test data is predominantly lacking. Progress has also been made in the development of non-animal test methods that can contribute to chemical sensitiser potency predictions (e.g. various peptide reactivity test methods; Gerberick et al. 2004; Schultz et al. 2005; Gerberick et al. 2007; Natsch and Gfeller 2008; Gerberick et al. 2009; Roberts et al. 2010) and in our understanding of how these data can be integrated to improve any potency prediction (Maxwell and Mackay 2008; Basketter and Kimber 2009; Natsch et al. 2009). However in the absence of any published evidence, the conclusion of this expert group is that these non-animal test data are not routinely applied to inform skin sensitisation risk assessment decisions. Therefore although progress has been made in the standardisation of non-animal test methods, this review concludes that the characterisation of sensitiser potency is currently not possible with sufficient confidence to allow risk assessment decisions to be made for the vast majority of cosmetic product exposure scenarios. Consequently, the most positive view of timing for full replacement (i.e. scientific ability to make skin sensitisation risk assessment decisions using a toolbox of non-animal test methods for all cosmetic ingredients and cosmetic product exposure scenarios) is 2017–2019. This timeline does not consider the time needed for validation and regulatory acceptance, but does consider the time required, typically 2–3 years, for demonstration that a non-animal test method is sufficiently robust and relevant to deliver information useful for the determination of the intrinsic relative potency of sensitising chemicals and so be of value for risk assessment decision-making. Furthermore although the timeline is based upon the premise that non-animal test methods predictive of each mechanistic step will be available, it is unlikely that information on every mechanistic step will be required to inform all risk assessment decisions. Therefore, it is expected that the scientific ability to inform skin sensitisation decisions without animal test data for *some* ingredients and exposure scenarios should be feasible ahead of 2017–2019.

Repeated dose toxicity

Executive summary

1. Repeated dose toxicity is present if a persistent or progressively deteriorating dysfunction of cells, organs

Table 2 Summary of identified alternative methods for skin sensitisation

Current endpoints addressed in animal test	Non-animal test methods	Part of sensitisation mechanism covered	Area(s) of applications	Status	Comments	Estimated time to have the method ready for entering pre-validation (years)
LLNA (OECD TG 429)	In silico alternatives (Q)SARs		Could become part of a toolbox for skin sensitisation hazard characterization	Optimised and in use to contribute to hazard identification.	See section "Mechanistic chemistry and in chemico reactivity (mechanistic step 2)"	N/A
Lymphocytes proliferation	Peptide reactivity	Haptenation	Could become part of a toolbox for skin sensitisation hazard characterization	Undergoing prevalidation at ECVAM (1 method)	See section "Mechanistic chemistry and in chemico reactivity (mechanistic step 2)"	0–1
GPMT				Optimised, not formally validated but extensive in-house evaluation (6+ methods)		
Buehler test (OECD TG 406)				Undergoing evaluation/prevalidation (1 method)	See section "Epidermal inflammatory responses (mechanistic step 3)"	0–2
Skin reaction	Keratinocytes cultures	Haptenation, skin inflammation	Could become part of a toolbox for skin sensitisation hazard characterization	Optimised, evaluation/prevalidation study conducted (1 method)		
				R&D (1+ methods)		
	Keratinocyte: dendritic cells co-cultures	Skin inflammation, DC activation	Could become part of a toolbox for skin sensitisation hazard characterization	R&D (1+ method)	See section "Keratinocyte/DC co-culture systems (mechanistic step 4)"	1–2
	3D reconstituted skin models	Skin Inflammation	Could become part of a toolbox for skin sensitisation hazard characterization	Undergoing evaluation/prevalidation (1 method)	See section "Human skin equivalents (reconstituted tissue models) (mechanistic step 3)"	0–1
	3D reconstituted skin models + dendritic cells	Skin Inflammation, DC activation	Could become part of a toolbox for skin sensitisation hazard characterization	R&D (1+ method)	See section "Human skin equivalents (reconstituted tissue models) (mechanistic step 3)"	2–5
	Dendritic cell activation	DC activation	Could become part of a toolbox for skin sensitisation hazard characterization	Undergoing prevalidation at ECVAM (2 methods)	See section "Dendritic cell responses (mechanistic step 4)"	0–2
				Undergoing evaluation/prevalidation (1 method)		
				Optimised (1+ methods)		
	Dendritic cell migration	DC migration	Could become part of a toolbox for skin sensitisation hazard characterization	R&D (1+ methods)	See section "Dendritic cell migration (mechanistic step 5)"	2–5
	T-cell proliferation	T-cell proliferation	Could become part of a toolbox for skin sensitisation hazard characterization	R&D (2+ methods)	See section "T cell responses (mechanistic step 6)"	5–7

This table estimates the time needed to enter prevalidation, according to ECVAM's established criteria, assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal test, nor does it include the time needed to achieve regulatory acceptance. "Optimal conditions" means that all necessary resources, for example, technical, financial, human and coordination, are met at all times in the process and that the studies undertaken have successful outcomes. None of the alternative methods available are suitable for quantitative risk assessment

- or multiple organ systems, results from long-term repeated exposure to a chemical. The onset and progression of this toxicity is influenced by the interplay between different cell types, tissues and organs, including the concomitant contribution of toxicokinetics, hormonal effects, autonomic nervous system, immuno- and other complex systems, which in certain cases are modulated by feedback mechanisms.
2. Since a wide range of endpoints is investigated in the animal repeated dose toxicity studies, an integrated approach based on the use of alternative methods with complementary endpoints needs to be developed.
 3. This chapter presents an overview of *in vitro* models in relation to six of the most common targets for repeated dose toxicity (liver, kidney, central nervous system, lung, cardiovascular and haematopoietic system). *In silico* tools such as (Q)SARs for predicting repeated dose toxicity are also discussed.
 4. The *in vitro* methods have been developed with the aim of producing stand-alone methods for predicting effects in specific target organs. Integrative research efforts considering interactions between different biological tissues and systems, which would be more representative of the situation *in vivo*, have only recently been initiated.
 5. Many of the identified tests are at the research and development level. The methods under development may be useful for hazard identification of target organ toxicity or for obtaining mechanistic information, but none of them is currently seen as appropriate for quantitative risk assessment for repeated dose toxicity. Prospective quantitative risk assessment for repeated dose toxicity by these methods is under development and will rely on the integration of biokinetic models.
 6. Intensive efforts will be necessary to optimise the existing models and to develop relevant *in vitro* models in those cases where fewer models are available.
 7. There is a need for more fundamental research focusing on understanding mechanisms of toxicity and toxicity pathways in support of predicting repeated dose toxicity, rather than on apical¹³ endpoints. Additional efforts are also necessary to develop improved biokinetic models to support extrapolation from *in vitro* to *in vivo* and understanding of dose response, in order for data obtained in *in vitro* models to be applied for quantitative risk assessment.
 8. Optimal use of existing data by the Threshold of Toxicological Concern (TTC) concept, read-across and integrated testing strategies can provide an opportunity to avoid the need for *in vivo* testing for a range of substances and applications. Work to further develop such approaches, incorporating information on toxicity pathways as it evolves, is recommended. This should include consideration of how information on consumer exposure can be used to determine the need for testing.
 9. In conclusion, the participating experts estimated that methods for full animal replacement with regulatory accepted tests/strategies will not be available by 2013. Full replacement for repeated dose toxicity is extremely challenging and the time needed to achieve this goal will depend on the progress at the research and development level and adequate prioritisation, funding and coordination of efforts.

Introduction

Repeated dose toxicity comprises the adverse general toxicological effects (excluding reproductive, genotoxic and carcinogenic effects) occurring as a result of repeated daily dosing with, or exposure to, a substance for a specified period up to the expected lifespan of the test species (ECHA 2008). Testing for repeated dose toxicity forms an integral part of the data package produced to perform quantitative risk assessment of cosmetic ingredients: the repeated dose study usually delivers the NOAEL (No observed adverse effect level) which is used in the calculation of the MoS (Margin of Safety) or MoE (Margin of Exposure).

Repeated dose testing *in vivo* permits observation of the integrated response to chemical exposure, including the concomitant contribution of toxicokinetics, hormonal effects, autonomic nervous system, immuno- and other complex systems, which in certain cases, are modulated by feedback mechanisms. In principle, it provides a relatively unbiased assessment in which all organs and toxicity endpoints (including those such as memory and behaviour) are covered and their relative importance evaluated. While differences between species mean that there is some uncertainty about the relevance of the results of such methods for predicting toxicity in humans, nevertheless, due to similarities in for example anatomy, biochemistry and physiology, in general repeated dose toxicity testing appears to be effective in safeguarding public health. Replacement of this *in vivo* methodology to evaluate such an integrated response (including compensatory responses), and in particular its quantitative aspects, is extremely challenging.

¹³ An observable outcome in a whole organism, such as a clinical sign or pathological state, that is indicative of a disease state that can result from exposure to a toxicant.

Current repeated dose toxicity methodology for the safety assessment of cosmetic ingredients

The following *in vivo* repeated dose toxicity tests are available for assessment of cosmetic ingredients:

1. Repeated dose (28 days) toxicity (oral) using rodents (EU 2008; OECD 2008a).
2. Repeated dose (28 days) toxicity (dermal) using rats, rabbits or guinea pigs (EU 2008; OECD 1981a).
3. Repeated dose (28 days) toxicity (inhalation) using rodents (EU 2008; OECD 2009a).
4. Sub-chronic oral toxicity test: repeated dose 90-day oral toxicity study in rodents (EU 2008; OECD 1998a).
5. Sub-chronic oral toxicity test: repeated dose 90-day oral toxicity study in non-rodents (EU 2008; OECD 1998b).
6. Sub-chronic dermal toxicity study: repeated dose 90-day dermal toxicity study using rodent species (EU 2008; OECD 1981b).
7. Sub-chronic inhalation toxicity study: repeated dose 90-day inhalation toxicity study using rodent species (EU 2008; OECD 2009e).
8. Chronic toxicity test using rodents (EU 2008; OECD 2009f).

The objective of repeated dose studies is to determine the potential toxic effects of a test substance in a mammalian species following daily administration for a prolonged period up to the whole lifespan of the animals and to yield the associated dose–response data. In these tests, effects that require a long latency period or are cumulative become manifest.

The most typical repeated dose studies are the 28- and 90-day studies (subchronic) in rats or mice. When required and justified, a chronic toxicity study lasting 52 weeks (1 year) may be conducted. In these studies, the test substance is administered in graduated doses to several groups of animals, one dose level per group, for the specified period (i.e. 28 days or 90 days). During the period of administration, animals are closely observed for signs of toxicity. Animals which die or are killed during the course of the test are necropsied and, at the conclusion of the test, surviving animals are also killed and necropsied. The dosing route (oral, dermal or inhalation) is usually selected depending on the potential human exposure route and the physico-chemical properties of the test substance (e.g. for highly volatile, easily evaporating substances the inhalation route is relevant).

From analysis of the Scientific Committee on Cosmetic products and Non-Food Products intended for consumers (SCC(NF)P) opinions over the last 10 years, the 90-day oral toxicity assay in rats is the most commonly conducted repeated dose toxicity study and the toxicity assay used

most to derive the NOAEL value for cosmetic ingredients (Pauwels et al. 2009; Van Calsteren 2010).

In principle, the repeated dose toxicity study yields the following data/information:

- general characteristics of the toxicity
- the target organs of toxicity
- the dose–response (curve) for each toxicity endpoint
- responses to toxic metabolites formed in the organism
- delayed responses, cumulative effects
- the margin between toxic/non-toxic dose
- NOAEL, NOEL for toxicity
- information on reversibility/irreversibility of the effect

Current availability and status of alternative methods for repeated dose toxicity

Introduction

In current OECD guidelines (OECD 1998a, b) for repeated dose toxicity testing in animals, in addition to gross examination, clinical signs, clinical chemistry and haematology, a comprehensive series of about 30 tissues and organs is examined histopathologically. These are the brain (representative regions including cerebrum, cerebellum and medulla/pons), spinal cord (at three levels: cervical, mid-thoracic and lumbar), pituitary, thyroid, parathyroid, thymus, oesophagus, salivary glands, stomach, small and large intestines (including Peyer's patches), liver, pancreas, kidneys, adrenals, spleen, heart, trachea and lungs, aorta, gonads, uterus, accessory sex organs, female mammary gland, prostate, urinary bladder, gall bladder (in mouse), lymph nodes (preferably one lymph node covering the route of administration and another one distant from the route of administration to cover systemic effects), peripheral nerve (sciatic or tibial) preferably in close proximity to the muscle, a section of bone marrow (and/or a fresh bone-marrow aspirate), skin and eyes (if ophthalmological examinations showed changes).

However, some tissues are much more frequently targets of toxicity than others. In a recent review on target organ involvement in attrition during drug development (Redfern et al. 2010), it was concluded that the cardiovascular system, nervous system and gastrointestinal system were most often involved. Other tissues that were affected on an appreciable number of occasions included the liver, immune system, respiratory system and musculoskeletal system. However, the relative importance of target organ varies with the stage of drug development. This is because some toxic reactions are idiosyncratic, so that they are detected only after market launch and exposure to large numbers of subjects. Hence, the targets most tractable to replacement of whole animals will be those identifiable

preclinically, as these will represent consistent, animal-based effects. Hence, the targets of most concern are the heart, the liver, the nervous system and the reproductive system.

While the kidney is not a major target in pre-clinical development of pharmaceuticals, it is clearly a frequent target for other chemicals, as is the lung (Bitsch et al. 2006). Hence, a toxicological assessment of repeated dose toxicity will need to include at least the liver, kidney, heart, lung and nervous system. Other targets of potential concern include the endocrine system, which covers many different organs and systems and will overlap with the reproductive system, the immune system, haematological system, including bone marrow, the musculoskeletal system and the gastrointestinal system. The skin and eye are also potential routes of exposure. For example, cataracts might not result from a single exposure to a chemical, but could be produced following repeat exposure to the eye. In addition, skin and eye can be potential target tissues following systemic exposure from other routes (e.g. oral, inhalation).

A survey of cosmetic ingredients, studied between 2006 and 2010 by the SCC(NF)P and Scientific Committee on Consumer Safety (SCCS) (Van Calsteren 2010), shows that repeated dose toxicity studies are present in nearly all dossiers. A 90-day study is most frequently reported (69% of cases). In 31%, a 28-day study was available, which was usually performed as a range finding study before a more elaborate 90-day study is conducted. Thus, 90-day sub-chronic toxicity studies usually deliver the NOAEL used in the calculation of the MoS.

The NOAEL value is in certain cases derived from teratogenicity studies (31%), but usually the 90-day repeated dose study provides the lowest value (in 77% of cases where both studies have been conducted), which is then used for safety reasons as the most conservative value for calculation of the MoS.

From this evaluation (Van Calsteren 2010; M. Pauwels and V. Rogiers, personal communication), it appeared that the most targeted organ was the liver, followed by the kidneys and spleen. Other organs targeted less frequently include the stomach, genitals, thyroid, adrenal glands, thymus and heart. Secondary parameters significantly changed were clinical biochemical values, haematology and body weight. Liver pathologies that could be tentatively connected with the changes observed were in particular steatosis and cholestasis.

Efforts to develop *in vitro* alternatives to animal testing have generally been based around predicting toxicity in a particular target organ, and these are discussed in the *in vitro* section below. However, the utility of such an approach is limited for quantitative risk assessment of repeated dose systemic toxicity, and it has been suggested

that approaches based on understanding of mode of action and biological pathways leading to toxicity may be of greater value (NRC 2007).

It should also be noted that as pharmaceuticals are subject to more extensive safety testing, and given the drive towards predictive toxicology in the pharma industry, most experience with non-animal methods has been gained with pharmaceuticals. For this reason, many of the methods discussed below have mainly been applied to pharmaceutical testing. Although only a limited set of assays may have been applied to cosmetics, most of the assays could potentially be used for this purpose.

(Q)SARs and in silico modelling

At present, only a few (Q)SAR models are available for repeated dose toxicity, the latter often being considered as a far too heterogeneous and complex endpoint to be encoded in a single predictive model. (Q)SARs based on general apical endpoints (toxicological effects), such as liver toxicity, are considered to have a low chance of success because of the diversity of mechanisms/modes of action involved in such effects. Hence, while (Q)SARs could have an important role to play, they may need to be more focused on specific mechanisms, and a suite of (Q)SARs will probably be necessary for most endpoints.

Nevertheless, initial attempts have suggested the feasibility of developing models providing meaningful predictions of chronic toxicity. An overview of commercially available models and those reported in the scientific literature is provided in Table 3.

It is worth also noting that most of these (Q)SARs have been developed to serve pharmaceutical needs. Although their applicability domains are wide and not specific to drugs, the sensitivity and specificity of these models are often uneven, and the extension to regulatory and cosmetics use should be done carefully. *In silico* models are best used in an integrated testing strategy and not as individual isolated tools. Examples of their validity and usability are given in the industrial strategies section.

Software: The provision of a numerical value (i.e. LOAEL; lowest observed adverse effect level) for potential use in quantitative risk assessment for repeated dose toxicity is currently only supported by the TOPKAT commercial package. The original model was developed by Mumtaz et al. (1995), and the updated and current module includes five regression (Q)SAR models based on 44 structural descriptors for five classes of chemicals (acyclics, alicyclics, heteroaromatics, single and multiple benzene rings) and was developed using 393 chemicals pulled together from various sources (US EPA and National Cancer Institute/National Toxicology Program (NCI/NTP) databases, FDA drug applications reports and the open

Table 3 Summary of in silico methods for repeated dose toxicity

Current endpoints addressed in animal test	Alternative tests available	Part of mechanism covered	Area(s) of application	Status (R&D, optimisation, prevalidation, validation, regulatory acceptance)	Comments	Estimated time until entry into pre-validation
[OECD 407, EC B.7] [OECD 408, EC B.26] [OECD 409, EC B.27] [OECD 410, EC B.9] [OECD 411, EC B.28] [OECD 412, EC B.8] [OECD 413, ECB 29] [OECD 452, EC B.30]	TOPKAT	Predicts LOAEL for chronic toxicity	Prioritisation/ screening	Optimised		No formal validation necessary, methods have to follow the OECD principles for the validation of QSARs for regulatory purposes
In all these animal tests the biological endpoints evaluated are clinical biochemistry, gross necropsy, histological evaluation and target organ	DEREK	Hepatotoxicity	Hazard identification	Optimised	Not quantitative—predicts potential hepatotoxicity (yes/no answer)	Formal validation not necessary
	DEREK	hERG channel inhibition (cardiotoxicity)	Hazard identification	Optimised	Not quantitative—predicts potential hERG inhibition (yes/no answer)	Formal validation not necessary
	LAZAR (Maunz and Helma 2008)	Predicts MRTD ^a for chronic toxicity	Prioritisation/ screening	Optimised	Developed using data on pharmaceuticals from clinical trials	Formal validation not necessary
	Garcia-Domenech et al. (2006)	Predicts LOAEL for chronic toxicity	Prioritisation/ screening	Optimised	Major work needed to develop software for wider use	Formal validation not necessary
	Mazzatorta et al. (2008)	Predicts LOAEL for chronic toxicity	Prioritisation/ screening	Optimised	Major work needed to develop software for wider use	Formal validation not necessary
	Matthews et al. (2004a, b)	Predicts MRTD ^a for chronic toxicity	Prioritisation/ screening	Optimised	Major work needed to develop software for wider use	Formal validation not necessary

^a Maximum recommended therapeutic dose

literature). The software was challenged by a number of independent studies (Venkatapathy et al. 2004; Tilaoui et al. 2007) showing that TOPKAT is able to predict approximately 30% of LOAELs within a factor of 2, 60% within a factor of 10 and 95% within a factor of 100. Although actual performances observed by these studies are dissimilar, the results should be considered in the light of the distinct differences in the data sets used.

Among its possible outputs, DEREK, another commercial software package, provides a prediction on hepatotoxicity potential. Experts have identified 74 structural alerts based on public-domain literature and proprietary data sets (Marchant et al. 2009). Validation results (positive predictivity of 80% for test chemicals which lie inside the applicability domain that is based on structural

fragments, 33% for chemicals which lie outside it) indicate that while these structural alerts are effective in identifying the hepatotoxicity of several chemicals, further research is needed to develop additional structural alerts to account for the hepatotoxicity of a number of chemicals that is not currently predicted. This model only flags hazardous structural alerts and thus is only applicable for the early identification of potential adverse liver effects after chemical exposure rather than for quantitative risk assessment.

Maunz and Helma (2008) have applied support vector regression (SVR) to predict the FDA MRTD (Maximum Recommended Therapeutic Dose) on the basis of local clusters of similar molecules. The MRTD is empirically derived from human clinical trials and is a direct measure

of the dose-related effects of pharmaceuticals in humans. In a pharmacological context, it is an estimated upper dose limit beyond which a drug's efficacy is not increased and/or undesirable adverse effects begin to outweigh beneficial effects. It is essentially equivalent to the NOAEL in humans, a dose beyond which adverse (toxicological) or undesirable pharmacological effects are observed. The SVR predictions of MRTD are obtained from the experimental results of compounds with similar structures (neighbours) with respect to the endpoint under investigation. Predictions are considered to be within the applicability domain when the confidence is lower than 0.2. 89% of predictions in the applicability domain are within 1 log unit from the experimental value. This performance drops to 82% when all predictions are considered. Authors have implemented this approach in the freely available Lazar software (<http://www.lazar.in-silico.de/>).

Models available in the scientific literature: Models published in peer-reviewed papers that could be useful for screening/prioritisation purposes if implemented in the form of software that also generates the necessary descriptors are listed below.

Garcia-Domenech et al. (2006) modelled the same data used in the TOPKAT training set (EPA and NTP reports) using graph theoretical descriptors and multilinear regression models for predicting the LOAEL in chronic studies. Although the models have different performances on the EPA pesticides database and the NTP database, they are transparent and have the advantage of being based on molecular connectivity indices, which are easily computed, invariant molecular descriptors. The error of the regression models was equivalent to the variance in the underlying experimental data. The results obtained should be considered in the light of the structural diversity of the training set.

Mazzatorta et al. (2008) reported a predictive *in silico* study of more than 400 compounds based on two-dimensional chemical descriptors and multivariate analysis. The training set included pesticides, drugs and natural products extracted from various reports (JECFA, JMPR, NCI and NIH) and the dataset of Munro et al. (1996a). It was found that the root mean squared error of the predictive model is close to the estimated variability of experimental values (0.73 vs. 0.64, respectively). The analysis of the model revealed that the chronic toxicity effects are driven by the bioavailability of the compound that constitutes a baseline effect plus excess toxicity possibly described by a few chemical moieties.

A model for predicting oral MRTD values in humans was developed by Matthews et al. (2004a). With the exception of chemotherapeutics and immunosuppressants, the MRTD/10 would correspond to a dose exerting neither therapeutic nor chronic adverse effects in humans

(Matthews et al. 2004b; Contrera et al. 2004). For non-pharmaceutical chemicals, there is no desired pharmacological effect and any compound-related effect could be interpreted as an adverse or non-desirable effect. The model classifies between high-toxicity chemicals and low-toxicity chemicals on the basis of structural alerts. While most of the training set has been identified (Matthews et al. 2004b), the algorithm is not provided. The model is reported to have a high positive predictivity and a low false positive rate, implying that it could be used to identify toxic chemicals.

In vitro models

Numerous *in vitro* systems have been generated over the last few decades, which claim to be useful either for predicting target organ toxicity or in assessing mechanistic aspects of target organ toxicity at the molecular, cellular and tissue level. The majority of the *in vitro* models are based on dispersed cell cultures, either as primary cultures or as continuous cell lines. The cells may be derived from many different species, including humans. The main drawbacks of primary cultures are their limited lifespan and that they do not always provide stable phenotypes. Cell lines may undergo dedifferentiation and lose the specific functional properties of the cells *in vivo* or express uncharacteristic functional features. In addition, *in vitro* cell culture systems often poorly resemble their *in vivo* equivalents, mainly due to the poor mimicking of the natural microenvironment. The use of extracellular support systems such as scaffolds and extracellular support matrix improves survival of cells in culture. Other approaches include the use of microporous supports (mono-cultures or co-cultures) in combination with perfusion systems to provide an organotypic environment that can improve differentiation and lifespan in culture (Prieto et al. 2006; Grindon et al. 2008). A substantial amount of research is currently being conducted on the potential use of stem cells as *in vitro* models for research and safety testing (Grindon et al. 2008; Chapin and Stedman 2009). Different types of differentiated cells can be obtained from embryonic and adult progenitor/stem cells of different species and also from induced pluripotent (iPS) cells (Banas et al. 2007; De Kock et al. 2009; Snykers et al. 2006, 2007; Taléns-Visconti et al. 2006).

Metabolism is an important issue for *in vitro* testing. It can be addressed to some extent by adding exogenous metabolising fractions or by using metabolically competent cells (Coecke et al. 2006; Combes et al. 2006; Bessems 2009). However, even if initially some metabolism is present, deterioration can occur as a function of culture time. Industry has now advanced new *in vitro* models to assess dermal penetration, including those of nanoparticles

and dermal metabolism (Jäckh et al. 2010; Landsiedel et al. 2010).

In the following section, a review of the *in vitro* models available for the six most common target organs for toxicity (see section “Introduction”) is provided. It should however be emphasised that there are many more target organs for which unfortunately less or no *in vitro* methods are available. None of the models can currently be applied for quantitative risk assessment for repeated dose toxicity.

Hepatotoxicity Because of its unique localisation and function in the organism, the liver, and the hepatocyte in particular, is a major target for toxicity. Hence, considerable attention has been paid over the years to the development of liver-based *in vitro* models. As a result of such efforts, a wide variety of hepatic *in vitro* systems is available for toxicity testing, ranging from subcellular hepatocyte fractions to whole isolated perfused livers (Elaut et al. 2006; Hewitt et al. 2007a). A prerequisite for repeated dose toxicity testing is that the method accurately and consistently predicts long-term effects. However, many of the hepatic cellular models undergo progressive changes in the functional and morphological phenotype, rendering them applicable for only short-term purposes. In the last decade, several innovative strategies have been introduced to counteract this dedifferentiation process, including genetic (Naiki et al. 2005) and epigenetic (Henkens et al. 2007) approaches in primary hepatocyte cultures. Similar strategies have been applied to liver-based cell lines, which are also prone to dedifferentiation (Martínez-Jiménez et al. 2006). A notable exception in this respect is the human hepatoma-derived HepaRG cell line, which persistently displays high functionality (Aninat et al. 2006). The maintenance of physiological functions over several weeks has been demonstrated for co-cultured human hepatocytes, which was developed based on a bioreactor technology for clinical bioartificial liver support (Schmelzer et al. 2009; Zeilinger et al. 2010).

While traditional high-order systems, such as precision-cut liver slices and isolated perfused livers, are still widely used for toxicity testing, there has been considerable focus in recent years on the development of bioartificial liver devices and perfused bioreactors consisting of primary hepatocytes or liver-based cell lines cultivated on micro-electronic sensors, micropatterned or microfluidic systems (Allen et al. 2005; Kim et al. 2008; Lee et al. 2007; Ma et al. 2009). A parallel track that has been followed in the last few years concerns the *in vitro* differentiation of hepatocyte-like cells from stem cell sources from different species (Guguen-Guillouzo et al. 2010; Snykers et al. 2009). This research area is still in its infancy, but given the exponentially growing interest, it can be anticipated that this field will be fully exploited in the upcoming years.

Significant progress has also been made lately with respect to the refinement of read-outs for toxicity testing in the available liver-based *in vitro* models, especially by combining them with “-omics” methodologies (Choi et al. 2010; De Gottardi et al. 2007; de Longueville et al. 2003; Elferink et al. 2008; Kikkawa et al. 2005, 2006; Li and Chan 2009; Meneses-Lorente et al. 2006; Mortishire-Smith et al. 2004; Petrak et al. 2006; Sawada et al. 2005). As part of the EU FP6 project Predictomics, an assay using a human liver cell line has been developed for the detection by flow cytometry of compounds with the potential to cause steatosis (Donato et al. 2009). An overview of the most commonly used *in vitro* tools for hepatotoxicity testing (i.e. measurement of apoptosis, necrosis, cholestasis, steatosis, phospholipidosis and fibrosis) is provided in Table 4. The potential use of these approaches for repeated dose toxicity testing is unclear because the interaction between different cells of the liver is hardly addressed, which might be of particular importance for assessing the capacity of the organ to regenerate after initial damage or for assessing adaptation processes.

Nephrotoxicity The kidney is a frequent target for many drugs and chemicals, some of which can contribute to end-stage renal diseases, which in Europe represent an important part of the total health-care burden. Moreover, the kidney appears to be the second most common target for cosmetic ingredients (Van Calsteren 2010). The kidney is a highly complex organ, composed of many different cell types, and it has a complex functional anatomy, so the assessment of normal or impaired organ function cannot easily be explored in *in vitro* studies (Prieto 2002).

Various *in vitro* models for the kidney are described in the literature. The most frequently used techniques such as renal slices, perfused nephron segments, isolated tubules as well as isolated tubular cells in suspension, longer-lasting primary cell cultures of isolated tubular cells and cell lines have been reviewed including their advantages and limitations (Boogaard et al. 1990; Pfaller et al. 2001). Among them, primary proximal tubular cells from rodents and humans are widely used as a renal cell model and many improvements in culture techniques have been reported (Hawksworth 2005; Li et al. 2006; Weiland et al. 2007; Wainford et al. 2009). In addition, a number of cell lines have been developed and used in nephrotoxicity testing. Most of these cell lines are derived from proximal tubular epithelium. They do not usually exhibit biotransformation activity, but they retain at least some capacity for the active transport of xenobiotics, which is relevant for kidney specific toxicity *in vivo*. Renal epithelial cell phenotypes with extensive glycolytic metabolism and morphology very close to the *in vivo* parent cell type are available, e.g. RPTEC/TERT1 cells, and can be maintained in culture up

Table 4 Alternative methods for repeated dose toxicity—hepatotoxicity

Current endpoints addressed in animal test	Alternative tests available	Part of mechanism covered	Area(s) of application	Status (R&D, optimisation, prevalidation, validation, regulatory acceptance)	Comments	Estimated time until entry into pre-validation (years)
[OECD 407, EC B.7]	Primary hepatocyte monolayer cultures (rodent/human)	Cytotoxicity (including necrosis and apoptosis), cholestasis, steatosis, phospholipidosis, fibrosis	Hazard identification (liver toxicants)	R&D	Innovative strategies to causally counteract phenotypic deterioration have been reported (e.g. epigenetic and genetic approaches)	>10
[OECD 408, EC B.26]					Significant progress has been made with respect to the development of “-omics”-based read-outs	
[OECD 409, EC B.27]					Various new cultivation scaffolds (natural or synthetic) have been introduced	>10
[OECD 410, EC B.9]	Primary hepatocyte sandwich cultures (rodent/human)	Cytotoxicity (including necrosis and apoptosis), cholestasis, fibrosis	Hazard identification (liver toxicants)	R&D		
[OECD 411, EC B.28]					A plethora of co-cultivation partners (hepatic or non-hepatic) have been described	>10
[OECD 412, EC B.8]	Primary hepatocyte co-cultures (rodent/human)	Cytotoxicity (including necrosis and apoptosis), fibrosis	Hazard identification (liver toxicants)	R&D		
[OECD 413, ECB 29]					The human hepatocellular carcinoma-derived HepaRG cell line shows high potential as an alternative test	<5
[OECD 452, EC B.30]	Cancer-derived or immortalised liver-based cell lines (rodent/human)	Cytotoxicity (including necrosis and apoptosis), cholestasis, steatosis, phospholipidosis, fibrosis	Hazard identification (liver toxicants)	R&D	Significant progress has been made with respect to the development of “-omics”-based read-outs	>10
In all these animal tests, the biological endpoints evaluated are clinical biochemistry, gross necropsy, histological evaluation and target organ					Significant progress has been made with respect to the development of “-omics”-based read-outs	>10
	Precision-cut liver slices (rodent/monkey/human)	Cytotoxicity (including necrosis and apoptosis), cholestasis, fibrosis	Hazard identification (liver toxicants)	R&D		
	Isolated perfused livers (rodent/human)	Cytotoxicity (including necrosis and apoptosis), cholestasis, steatosis, fibrosis	Hazard identification (liver toxicants)	R&D		
	Primary hepatocytes or liver-based cell lines in 96-well plate screening system	Cytotoxicity (including necrosis and apoptosis), phospholipidosis	Hazard identification (liver toxicants)	R&D		
	Primary hepatocytes or liver-based cell lines cultivated on microelectronic sensors, micropatterned devices or microfluidic systems	Cytotoxicity (including necrosis and apoptosis)	Hazard identification (liver toxicants)	R&D		
	Perfused bioreactors containing primary hepatocytes, liver-based cell lines or mixed cell populations in various configurations	Cytotoxicity (including necrosis and apoptosis)	Hazard identification (liver toxicants)	R&D		
	Hepatocyte-like cells derived from embryonic stem cells	Cytotoxicity (including necrosis and apoptosis)	Hazard identification (liver toxicants)	R&D	Assays using hepatocyte-like cells differentiated from other stem cell sources (e.g. adult tissue) are in the pipeline	>10

This table estimates the time needed to enter prevalidation, according to ECVAM's established criteria, assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal test, nor does it include the time needed to achieve regulatory acceptance. “Optimal conditions” means that all necessary resources, for example, technical, financial, human and coordination, are met at all times in the process and that the studies undertaken have successful outcomes. None of the alternative methods available are suitable for quantitative risk assessment

Table 5 Alternative methods for repeated dose toxicity—nephrotoxicity

Current endpoints addressed in animal test	Alternative tests available	Part of mechanism covered	Area(s) of application	Status (R&D, optimisation, prevalidation, validation, regulatory acceptance)	Comments	Estimated time until entry into pre-validation (years)
[OECD 407, EC B.7] [OECD 408, EC B.26] [OECD 409, EC B.27] [OECD 410, EC B.9] [OECD 411, EC B.28] [OECD 412, EC B.8] [OECD 413, ECB 29] [OECD 452, EC B.30]	Primary rat kidney tubular epithelial cells	Nephrotoxicity	Hazard identification	R&D		>10
In all these animal tests, the biological endpoints evaluated are clinical biochemistry, gross necropsy, histological evaluation and target organ	Primary human kidney tubular epithelial cells	Nephrotoxicity	Hazard identification	R&D	Limited availability	>10
	Human proximal tubule epithelial cell lines, e.g. RPTEC/TERT1, HK-2, HEK-293	Nephrotoxicity	Hazard identification	R&D	RPTEC is most promising cell line	5–10
	Rat proximal tubule epithelial-like cell line	Nephrotoxicity	Hazard identification	R&D		>10
	Pig proximal tubule epithelial cell line	Nephrotoxicity	Hazard identification	R&D		>10
	Dog distal tubule cell line	Nephrotoxicity	Hazard identification	R&D		>10

This table estimates the time needed to enter prevalidation, according to ECVAM's established criteria, assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal test, nor does it include the time needed to achieve regulatory acceptance. "Optimal conditions" means that all necessary resources, for example, technical, financial, human and coordination, are met at all times in the process and that the studies undertaken have successful outcomes. None of the alternative methods available are suitable for quantitative risk assessment

to 6 weeks (Wieser et al. 2008; Aschauer et al. 2010; Crean et al. 2010). The usefulness of RPTEC/TERT1 cells to study nephrotoxicity is currently under evaluation in the EU FP7 project Predict-IV. Critical pathways have been identified as the antioxidant and detoxification Nrf2 pathway in human HK-2 cells (Jennings 2010; Wilmes et al. 2010).

In vitro models for other segments of the nephron are limited. No model is available for assessment of the potential for toxicity to the kidney medulla. An overview of the many available models is given in Table 5.

Cardiovascular toxicity As noted previously, the cardiovascular system is one of the most commonly affected targets associated with attrition during drug development and is also the most common cause of drug withdrawal from the market (Kettenhofen and Bohlen 2008; Stummann et al. 2009a). A number of in vitro assays have therefore been developed for the screening of new pharmaceuticals for potential cardiotoxicity. The heart may also be target of cosmetic ingredients albeit to a lesser extent than several other organs (Van Calsteren 2010). The assays developed for testing of pharmaceuticals may therefore also have potential for the assessment of cardiac effects

from chronic repeated exposures to cosmetic ingredients. An overview of available in vitro models is provided in Table 6.

Many of these assays focus on detection of the ability to block the hERG potassium channel, which causes QT prolongation, including receptor binding assays, ion efflux assays and patch clamp studies in cell lines expressing the hERG channel (Houck and Kavlock 2008).

Primary cardiomyocytes have been isolated from a variety of animal species and have also been used in toxicological studies. An advantage of these cells is that they express all the ion channels underlying the cardiac action potential, but adult cells have a low proliferative capacity. Foetal and neonatal cardiomyocyte isolations may also contain other cell types such as fibroblasts, which will overgrow the cardiomyocytes after a few days in culture (Kettenhofen and Bohlen 2008). Recently, new culture techniques for the long-term culture of primary neonatal mouse cardiomyocytes have been described (Sreejit et al. 2008).

Cardiomyocytes have also been derived from embryonic stem (ES) cells. To assess cardiac-specific toxicity, the effects of compounds in ES cell-derived cardiomyocytes can be compared to those in non-cardiac cells such as

Table 6 Alternative methods for repeated dose toxicity—cardiotoxicity

Current endpoints addressed in animal test	Alternative tests available	Part of mechanism covered	Area(s) of application	Status (R&D, optimisation, prevalidation, validation, regulatory acceptance)	Comments	Estimated time until entry into pre-validation
[OECD 407, EC B.7] [OECD 408, EC B.26] [OECD 409, EC B.27] [OECD 410, EC B.9] [OECD 411, EC B.28] [OECD 412, EC B.8] [OECD 413, ECB 29] [OECD 452, EC B.30]	Murine embryonic stem cell-derived cardiomyocytes	Cytotoxicity Alterations in cardiac action potential	Screening/ hazard identification	R&D		>10 years
In all these animal tests, the biological endpoints evaluated are clinical biochemistry, gross necropsy, histological evaluation and target organ	Human embryonic stem cell-derived cardiomyocytes	Cytotoxicity Alterations in cardiac action potential	Screening/ hazard identification	R&D	Differentiation endpoints not clearly defined resulting in a variation of cell types and developmental status of the cells.	>10 years
	Primary cardiomyocytes from a variety of animal species	Cytotoxicity Alterations in cardiac action potential	Screening/ hazard identification	R&D	Disadvantage of adult primary cardiac cells is their low proliferative capacity.	>10 years
	Cell lines	Effects on ion channels—primarily hERG	Screening/ hazard identification	hERG assays used in pharmaceutical development	Multichannel effects are not properly detected.	Ready
	Engineered cardiac tissue	Cytotoxicity Arrhythmia	Screening/ hazard identification	Commercial systems for drug screening available		Ready

This table estimates the time needed to enter prevalidation, according to ECVAM's established criteria, assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal test, nor does it include the time needed to achieve regulatory acceptance. "Optimal conditions" means that all necessary resources, for example, technical, financial, human and coordination, are met at all times in the process and that the studies undertaken have successful outcomes. None of the alternative methods available are suitable for quantitative risk assessment

fibroblasts. Biomarkers of cardiac damage are released by ES cell-derived cardiomyocytes, and electrophysiological behaviour can also be assessed (Kettenhofen and Bohlen 2008).

Efforts are also ongoing to develop three-dimensional tissue engineering models of cardiac tissue, for both clinical replacement and toxicity assessment purposes (Kettenhofen and Bohlen 2008; Franchini et al. 2007). Engineered heart tissue models derived from neonatal rat cardiac myocytes and human ES cells are commercially available, and it is claimed that these can be used to screen for arrhythmogenic and cardiotoxic effects of pharmaceuticals (<http://www.tristargroup.us/>).

Overall, it can be concluded that alternative tests in the field of drug-induced arrhythmia are in a relatively advanced stage, but are mainly performed in addition to in vivo studies to improve drug safety. They may also be of relevance for testing of cosmetic ingredients, but primarily in the context of hazard identification. Fewer alternative

methods are currently available to evaluate the potential for compounds to cause contractility toxicity, ischaemic effects, secondary cardiotoxicity and valve toxicity (Stummann et al. 2009a). Pressurised human arteries have been used for the investigation of microvascular dysfunction including vascular permeability and flow-mediated dilatation (Moss et al. 2010), but methods sufficient for quantitative risk assessment of repeated dose cardiac toxicity are not yet available.

Neurotoxicity Detection of neurotoxicity induced by chemicals represents a major challenge due to the physiological and morphological complexity of the central (CNS) and peripheral nervous system (PNS). Neurotoxicity is currently evaluated during repeated dose toxicity studies using in vivo methods that are based mainly on the determination of neurobehavioural and neuropathological effects. At present, validated in vitro methods for neurotoxicity that can provide quantitative predictions for use in

risk assessment are not available. There are in vitro methods which are used for screening purposes and to improve mechanistic understanding of processes underlying normal or pathological nervous system function.

The central nervous system is comprised of various cell types (neuronal and glial), complex cell–cell interactions and unique protein interactions where functional coupling via synapses, gap junctions, signalling molecules and growth factors has to be preserved.

Several in vitro systems from single cell types to systems that preserve some aspects of tissue structure and function are currently available for toxicity testing. The available models consist of primary culture, neuroblastoma and glioma cell lines and recently available neural stem cell lines (both human and rodent) (Bal-Price et al. 2008; Harry et al. 1998). In the case of peripheral nervous system (the ganglia and the peripheral nerves lying outside of the brain and spinal cord), also various cell culture models exist including primary culture and cell lines; however, only limited aspects of Schwann cell function (interaction with axons) can be examined (Harry et al. 1998; Suuronen et al. 2004; Moore et al. 2005). Botulinum neurotoxin (BoNT) potency testing does not fall under cosmetic regulations, and therefore, it is not discussed here. The current scientific and legal status of alternative methods to the LD50 test for botulinum neurotoxin potency testing has been reported by Adler et al. (2010).

In vitro models of the blood–brain barrier (BBB) are also available, as it is necessary to define whether a compound crosses the BBB and whether it induces a direct toxic effect on the BBB (Cecchelli et al. 2007).

These in vitro approaches allow the assessment of cell viability, general but critical cell functions such as energy metabolism, oxidative stress and calcium homeostasis, and neuronal specific functions (neurite outgrowth and axonal transport, synaptogenesis/myelination, neurotransmission and vesicular release, signalling between neurons and glia, receptor pharmacology, ion channel activation, electrical activity, etc.). To date, in vitro neurotoxicity testing has been used mainly for mechanistic studies, where molecular/cellular pathways of toxicity are determined and used as the readout of chemically induced neuronal and glial damage.

In vitro models for neurotoxicity testing:

1. Dissociated primary cell culture monolayer (mixed: neuronal/glial) is the most widely used in vitro system for neurotoxicity evaluation. It allows for visualisation of individual living cells (neuronal and glial) and monitoring both morphological and electrophysiological features. Dissociated cell cultures are more accessible and easier to obtain and maintain. Additional purification methods can be used to enrich a

particular cell type (neuronal or glial). However, in vivo–like structures cannot be achieved by this technique. This model is currently under evaluation by “omics” technology and electrical activity measurement in the EU FP7 Predict-IV project.

2. Reaggregate cultures (or explant/slice culture) offer a more structured, three-dimensional histotypic organisation so more closely approximate in vivo conditions for cell growth and development. Processes such as synaptogenesis/myelination, neurotransmission and vesicular release are the most classical endpoints studied in 3D models. This model is currently under evaluation by “omics” technology in the EU FP7 Predict-IV project.
3. Continuous cell lines of tumoural origin provide homogeneous cell populations (neuronal or glial) in large quantities in a very reproducible manner. However, neuronal–glial cell interaction is lost. Many of these cell lines display properties of their normal cell counterpart when a differentiated state is induced by NGF, retinoic acid, etc.

Recently, different types of stem cells (either adult, embryonic or those derived from cord blood) are used as a source of neural progenitor cells that can be differentiated into functional neuronal and glial cells with possible applications for neurotoxicity testing (Buzanska et al. 2009). The advantage of this approach is that it is a human model and can be maintained in culture at different developmental stages: as non-differentiated stem cells, committed progenitors and lineage directed into neuronal, astrocytic and oligodendroglial cells. Based on such models, a complex in vitro testing strategy has to be developed with a battery of complementary endpoints using high-throughput and/or high content screening platforms to be able to test large number of substances with different mechanisms of toxicity. However, before these tests are used for routine screening, the sensitivity, specificity and reliability of the endpoints and models, and their capacity to predict human neurotoxic effects should be established.

An overview of available in vitro models is provided in Table 7. The in vitro models, developed so far for neurotoxicity, are limited to screening, prioritisation and hazard identification. They are not suitable for quantitative risk assessment.

Pulmonary toxicity The respiratory tract consists of different cell populations including epithelial, nervous, endothelial and immune cells which may respond to a variety of stimuli including inhaled chemicals. Damage may cause airway obstruction, inflammation or hyperresponsiveness, resulting in a variety of lung diseases which are a major global health concern. Chronic obstructive

Table 7 Alternative methods for repeated dose toxicity—neurotoxicity

Current endpoints addressed in animal test	Alternative tests available	Part of mechanism covered	Area(s) of application	Status (R&D, pre-validation, regulatory acceptance)	Comments	Estimated time until entry into pre-validation
[OECD 407, EC B.7]	Micro-electrode arrays (MEA) for single cell recordings exist for both evoked and spontaneous activity.	Relevant for neurophysiology testing	Screening/prioritisation	Optimisation and standardisation (ongoing ring trial with three European and two USA labs)	Very promising endpoint (sensitive and neuronal specific, based on the network function)	2011
[OECD 408, EC B.26]	Suitable in vitro models for MEA are primary rodent culture of cortex, hippocampus, spinal cord, etc.	Disturbed neuronal electrical communication in the network			Some promising results also obtained with human neural progenitor cells differentiated into neurons	
[OECD 409, EC B.27]		It allows to distinguish between neurotoxicity and cytotoxicity				
[OECD 410, EC B.9]						
[OECD 411, EC B.28]						
[OECD 412, EC B.8]						
[OECD 413, ECB 29]	Specific receptors and channels function (both excitatory and inhibitory): e.g. NMDA, GABA, in primary cultures	Neurotransmission	Hazard identification	Developed		>3 years
[OECD 452, EC B.30]		Changes in the populations of various receptors (up or down regulation as well as conductance through different channels)				
In all these animal tests, the biological endpoints evaluated are clinical biochemistry, gross necropsy, histological evaluation and target organ	Measurement of NTE (neurotoxic esterase) and AChE activity in SH-SY5Y cell line differentiated into neuronal phenotype (e.g. organophosphates in the presence of retinoic acid)	Enzymatic activity	Screening/hazard identification	Optimised	ECVAM received in 2010 a full submission for validation studies after positive evaluation of the pre-submission form	2011
If specific symptoms of neurotoxicity are observed the OECD TG424 has to be applied	Measurement of enzymatic activity specific for each cell type (neuronal and glial) in existing primary in vitro models (e.g. dissociated primary cell culture monolayer, reaggregate cultures)	Enzymatic activity	Hazard identification	R & D		>5 years
	All existing in vitro models (e.g. dissociated primary cell culture monolayer, reaggregate cultures, continuous cell lines)	Relevant for general biochemistry	Hazard identification	Developed and standardised assays	Neurochemistry can be covered by in vitro approach to a large extent	Most of them could enter prevalidation
	Non-mammalian models, e.g.: zebra fish, medaka or <i>C. elegans</i>	Neurotoxicity induced by changes in general but critical cell process measured by gene and protein expression, energy metabolism, oxidative and nitrosative stress	Hazard identification and screening	Some behaviour tests in non-mammalian models are well optimised and standardised	In vitro methods are not relevant to behaviour studies. Therefore, at present the only possibility is to use alternative non-mammalian models	Some of them ready for prevalidation
	Primary culture (rodent or human) monolayer or 3D (re-aggregating)	Neurobehaviour: Various (extremely complex mechanisms involved)	Hazard identification	R&D		>5 years
	Human and rodent stem cells differentiated into neuronal or glial cell type	Cell morphology/differentiation state				
		Cell differentiation evaluated by neurite outgrowth, number of processes per cell body				
		Evaluation of the cell population (number of neuronal types and glial in each relevant model) by protein expression specific for each cell type (GFAP for astrocytes, NF-68 or NF-200, β Tubulin-III for neurons, MBP for oligodendrocytes)				
	Differentiated neuroblastoma cell line (e.g. SH-SY5Y) or PC12 in the presence of NGF or retinoic acid	Cell morphology/differentiation state	Hazard identification and screening	Optimised	Cell line in particular are suitable models for neurite outgrowth measurements (e.g. PC12 cells)	Some ready for prevalidation

This table estimates the time needed to enter pre-validation, according to ECVAM's established criteria, assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal test, nor does it include the time needed to achieve regulatory acceptance. "Optimal conditions" means that all necessary resources, for example, technical, financial, human and coordination, are met at all times in the process, and that the studies undertaken have successful outcomes. None of the alternative methods available are suitable for quantitative risk assessment

pulmonary disease (COPD) is becoming the third leading cause of death worldwide (WHO 2008). Asthma is currently the most frequent chronic disease affecting children. No validated or widely accepted *in vitro* methods are yet available for the identification and characterisation of chemicals that have the potential to cause lung toxicity (Bérubé et al. 2009). Different *in vitro* models are being developed based on primary cell cultures derived from human (lung slices, biopsies, bronchioalveolar lavage) or rodent tissues. Some of the systems maintain an organotypic structure up to 12 months. Lung disease models (COPD, asthma, smoker, cystic fibrosis) and devices for long-term exposure of epithelia to vapours and aerosols are also available (Hayden et al. 2010) (<http://www.mattek.com>). For the safety assessment of nanomaterials, organotypic lung models have been established to study a possible inflammatory response (Brandenberger et al. 2010). Epithelia can be stimulated regularly with pro-inflammatory substances to simulate chronic inflammatory reactions for up to several months. Organotypic lung models exhibit an *in vivo*-like expression pattern of CYP450 enzymes (Constant 2010; <http://www.epithelix.com>). An organotypic lung model suitable for long-term toxicity testing is under evaluation (Huang 2009). However, these models have not yet been applied or validated for quantitative risk assessment.

Cell lines are available with characteristics that mimic different cell types of the respiratory tract. Co-cultures are also being developed. In addition, epithelial airway cells can be cultured at the air–liquid interface. Devices are being developed that represent the *in vivo* respiratory air compartment and which allow exposure of the cells to gases, liquid aerosols, complex mixtures, nanoparticles and fibres (Deschl et al. 2010; Gminski et al. 2010). This implies that cosmetics can be applied to potential target cells in realistic conditions, and exposure can be accurately monitored. An overview of available *in vitro* models is provided in Table 8.

The endpoints that can be measured include cytotoxicity, morphology, cilia beating, mucus secretion, damage of specific cell functions such as expression of cytokines, chemokines, metalloproteinases, lung specific molecules and changes in gene expression. Most of the models are used for mechanistic studies, and no formal validation study has been conducted yet.

Immunotoxicity and myelotoxicity Immunotoxicity is defined as the toxicological effects of xenobiotics on the functioning of the immune system and can be induced in either direct (caused by the effects of chemicals on the immune system) or indirect ways (caused by specific immune responses to the compounds themselves or to self-antigens altered by these compounds) (Lankveld et al. 2010).

At present, immunotoxicity is evaluated mainly *in vivo*. The OECD test guideline No. 407 includes parameters of immunotoxicological relevance as part of a repeated dose 28-day oral standard toxicity study in rodents (OECD 2008a). This guideline indicates that information about the toxicity on the immune system can be obtained through the analysis of total and absolute differential leucocyte counts, detection of globulin serum level, gross pathology and histology of lymphoid organs, organ weight (of thymus and spleen), and histology of bone marrow, BALT (bronchus-associated lymphoid tissues) and NALT (nasal-associated lymphoid tissues).

The general view is that effects on the immune system are very difficult to reproduce *in vitro*, because of the requirement of complex cellular interactions. However, some isolated processes may be studied *in vitro* such as proliferation of T lymphocytes and cytokines release (Carfí et al. 2010).

In vitro models could be used for pre-screening of immunotoxic potential (i.e. hazard identification), as part of a strategy (Carfí et al. 2007). The first tier will consist of measuring myelotoxicity. At present, a scientifically validated human and murine *in vitro* colony-forming unit-granulocyte/macrophage (CFU-GM) assay is available for evaluating the potential myelotoxicity of xenobiotics (Pessina et al. 2003; Pessina et al. 2010). Toxic effects on proliferation and differentiation of progenitor cells of different blood cell lineages can also be measured *in vitro* (Pessina et al. 2005), as well as long-term repopulating capacity of more primitive haemopoietic stem cells and the stromal microenvironment (Broxmeyer et al. 2006; Miller and Eaves 2002; Podestà et al. 2001; Sutherland et al. 1990). If the compound is not myelotoxic, it can be tested for lymphotoxicity in the second tier. Several *in vitro* assays for lymphotoxicity exist, each comprising specific functions of the immune system, cytokine production (Langezaal et al. 2002; Ringerike et al. 2005), B- and T-cell proliferation (Carfí et al. 2007; Smialowicz 1995), cytotoxic T-cell activity (House and Thomas 1995), natural killer cell activity (Kane et al. 1996), antibody production and dendritic cell maturation (Mellman and Steinman 2001).

An overview of available *in vitro* models is provided in Table 9.

Omics and imaging technologies

In recent years, transcriptomics (i.e. whole genome gene expression analysis based on microarray technology) has been applied to *in vitro* models of human and rodent cells for the purpose of predicting toxicity, for example, with respect to genotoxicity/carcinogenicity, target organ toxicity and endocrine disruption.

Table 8 Alternative methods for repeated dose toxicity—pulmonary toxicity

Current endpoints addressed in animal test	Alternative tests available	Part of mechanism covered	Area(s) of application	Status (R&D, optimisation, prevalidation, validation, regulatory acceptance)	Comments	Estimated time until entry into pre-validation (years)
[OECD 407, EC B.7]	<i>Primary cultures</i>					
[OECD 408, EC B.26]	Normal human bronchial/nasal epithelial cells	Allergic asthma	Hazard identification	R&D		>10
[OECD 409, EC B.27]	Human airway gland epithelium	Chronic obstructive airway diseases	Hazard identification	R&D		>10
[OECD 410, EC B.9]		Cystic fibrosis				
[OECD 411, EC B.28]		Adaptive immunity	Hazard identification	R&D		>10
[OECD 412, EC B.8]	Dendritic cells (CD) derived from human peripheral blood (monocyte derived) or CD34+ cord blood stem cells	Allergic asthma				
[OECD 413, ECB 29]						
[OECD 452, EC B.30]						
In all these animal tests the biological endpoints evaluated are: clinical biochemistry, gross necropsy, histological evaluation and target organ	3D-long-term human cell cultures of nasal/bronchial/alveolar epithelial cells Kept for up to 12 months	Inflammation (asthma) Remodelling (COPD) Mucus secretion (Cystic fibrosis) Ciliary clearance Active ion transport	Hazard identification	R&D		>5
	Lung slices	Immunomodulation	Hazard identification	R&D		>10
	<i>Cell lines</i>					
	Bronchial epithelial cell lines (e.g. BEAS-2B, Calu-3, 16HBE14o_)	Chronic obstructive airway diseases Allergic asthma	Hazard identification	R&D		>10
	Alveolar epithelial cell lines, e.g. A549	Inflammation irritation	Hazard identification	R&D		>10
	Macrophage cell lines THP-1 and U-937	Inflammation	Hazard identification	R&D		>10
	Human DC-like cell lines, e.g. KG-1, MUTZ-3, U-937 and THP-1	Sensitisation	Hazard identification	R&D		>10
	3D cultures of A549 alveolar macrophages	Irritation	Hazard identification	R&D		>10
	Co-culture models Epithelial cells with dendritic cells Alveolar macrophages and human bronchial epithelial cells	Inflammation	Hazard identification	R&D		>10

This table estimates the time needed to enter prevalidation, according to ECVAM's established criteria, assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal test, nor does it include the time needed to achieve regulatory acceptance. "Optimal conditions" means that all necessary resources, for example, technical, financial, human and coordination, are met at all times in the process, and that the studies undertaken have successful outcomes. None of the alternative methods available are suitable for quantitative risk assessment

Table 9 Alternative methods for repeated dose toxicity—immunotoxicity and myelotoxicity

Current endpoints addressed in animal test	Alternative tests available	Part of mechanism covered	Area(s) of application	Status (R&D, optimisation, prevalidation, validation, regulatory acceptance)	Comments	Estimated time until entry into pre-validation
	<i>Immunotoxicity</i>					
[OECD 407, EC B.7]	Human whole-blood cytokine release assay	TH2 responses	Hazard identification	Prevalidated		Prevalidated
[OECD 408, EC B.26]	Lymphocyte proliferation assay	Adaptive immune responses	Hazard identification	In validation		In validation
[OECD 409, EC B.27]	On human peripheral blood lymphocytes					
[OECD 410, EC B.9]	Murine spleen cells					
[OECD 411, EC B.28]	Mixed lymphocyte reaction (rodent lymphocytes)	Response to allogenic cells	Hazard identification	R&D		>10 years
[OECD 412, EC B.8]	Cytotoxic T-lymphocyte assay (rodent splenocytes/human peripheral blood mononuclear cells)	Differentiation of T cells into cytotoxic effector cells	Hazard identification	R&D		>10 years
[OECD 413, ECB 29]	Natural Killer cell assay	Nonspecific defence system against tumours, microbe-infected cells, and MHC Class I mismatched cells	Hazard identification	R&D		>10 years
[OECD 452, EC B.30]	Human peripheral blood cells	Adaptive immune responses	Hazard identification	R&D		>10 years
In all these animal tests the biological endpoints evaluated are: clinical biochemistry, gross necropsy, histological evaluation and target organ	<i>Myelotoxicity</i>					
	Fluorescent cell chip using transgenic cell lines	Immunomodulation	Hazard identification	R&D		>10 years
	Human/murine/rat CFU-GM	Proliferation capacity and maturation of progenitor cells	Hazard identification	Scientifically validated		
	CFU-E (colony-forming unit-erythroid)	Proliferation capacity and maturation of progenitor cells	Hazard identification	R&D		>10 years
	BFU-E (burst-forming unit-erythroid)					
	CFU-G (CFU-granulocyte)					
	CFU-M (CFU-macrophage)					
	CFU-GEMM (CFU-granulocyte/erythroid/macrophage/megakaryocyte)					
	Human or rodent					
	Long-term bone-marrow cultures (LTBC)	Proliferation capacity of primitive haemopoietic cells	Hazard identification	R&D		>10 years
	Human or rodent	Quality of the stromal microenvironment				
	LTC-initiating cell (LTC-IC)	Differentiation of primitive haemopoietic cells	Hazard identification	R&D		>10 years
	Human or rodent					
	Cobblestone area forming cell (CAFC) assay	Differentiation of primitive haemopoietic cells	Hazard identification	R&D		>10 years
	Human or rodent					

This table estimates the time needed to enter prevalidation, according to ECVAM's established criteria, assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal test, nor does it include the time needed to achieve regulatory acceptance. "Optimal conditions" means that all necessary resources, for example, technical, financial, human and coordination, are met at all times in the process, and that the studies undertaken have successful outcomes. None of the alternative methods available are suitable for quantitative risk assessment.

In developing transcriptomics-based screens for toxic class prediction by using DNA microarray technology, gene expression data are derived from exposure of model systems (such as cellular models) to well-known toxicants belonging to specified classes of toxicity. From these data sets, advanced statistics can be applied to identify generic gene expression profiles corresponding to different types of toxicity. These profiles are compared to a set of gene expression changes elicited by a suspected toxicant. If the characteristics match, a putative mechanism of action can be assigned to the unknown agent and hazard predicted.

Several investigations have demonstrated that predictive transcriptomic profiling may be achievable in vivo after short-term treatment periods, thus showing its potential for informing repeated dose toxicity testing in rodent models and possibly using in vitro systems (Steiner et al. 2004; Fielden et al. 2007; Kier et al. 2004; Jiang et al. 2007).

Several studies have been published which used 'omics techniques to determine the value of conventional or optimised in vitro hepatocyte models for toxicogenomics investigations. A few examples are highlighted below. Boess et al. (2003) compared base-line gene expression profiles between primary hepatocytes either cultured conventionally or in a collagen-sandwich culture, liver slices and immortal cell lines of liver origin with gene expression profiles in the in vivo liver, all from rat. They found liver slices exhibited the strongest similarity to liver tissue regarding mRNA expression, whereas the two cell lines were quite different from the whole liver. Kienhuis et al. (2007) demonstrated by means of gene expression profiling in combination with enzyme activity assays that a modified rat hepatocyte-based in vitro system enriched with low concentrations of well-known enzyme inducers offers an improvement over existing models with respect to sustaining metabolic competence in vitro. An inter-laboratory comparison of transcriptomic data obtained from a human proximal tubule cell line (HK-2) found that the microarray data were generally satisfactory, although confounding factors such as medium exhaustion during cell culture must be considered (Jennings et al. 2009).

Because of our strongly increased insights into the complexity of cellular biology, current efforts in developing in vitro omics-based alternatives for animal models for repeated dose toxicity aim at understanding mechanisms of action by in-depth investigations of molecular response pathways. Transcriptomics is not the only technology that is being employed for such research; microRNA analysis, epigenetics, proteomics and metabonomics all have an important role to play. In addition, novel imaging technologies and physiological analyses such as impedance measurements provide the possibility for continuous observation of major cellular events such as migration,

proliferation, cell morphology, cell–cell interactions and colony formation.

As an example, the EU FP7 project Predict-IV is evaluating the integration of 'omic technologies, biomarkers and high content imaging for the early prediction of toxicity of pharmaceuticals in vitro. The aim is to identify general pathways resulting in toxicity that are independent of the cell/tissue type.

While these technologies hold considerable promise, further development is needed before they can be applied for quantitative risk assessment for repeated dose toxicity. Stable and reproducible long-term culture systems are needed before omics provide data that actually can be used.

Strategies to reduce, refine or replace the use of animals

Although no full replacement alternatives for in vivo studies with regard to quantitative risk assessment for repeated dose toxicity are likely to become available for some time, alternative methods and use of integrated testing strategies should be used to refine and reduce the use of laboratory animals. Reduction/refinement can be achieved with regard to species selection, selection of the most relevant route, toxicokinetically guided study design (Barton et al. 2006; Creton et al. 2009), better selection of doses, use of intermediate endpoints, assessment of urgency (priority setting) and waiving of the need to perform in vivo bioassays in certain circumstances (Bessemers 2009; Vermeire et al. 2010). In the case of cosmetic ingredients, human systemic exposure may be estimated on the basis of in vitro skin penetration data; however, the limits of this test should also be taken into account to avoid over-predicting exposure and therefore avoid triggering unnecessary in vivo testing (Nohynek et al. 2010).

Positive predictions based on a set of in vitro assays could be used to focus planning of in vivo assays and use less animals (for confirmation) or finish the study at an earlier time-point by using more specific and/or earlier and more sensitive endpoints or biomarkers. Further use of toxicokinetic data in in vitro approaches can also decrease the number of animals needed for confirmation of in vivo findings.

Integrated testing strategies Integration of multiple approaches and methods has greater potential to provide relevant data for a weight of evidence approach in safety evaluation and risk assessment than individual tests. Intelligent testing strategies may include combinations of chemical category, read-across, (Q)SAR, in vitro and in vivo methods (Bessemers 2009; Schaafsma et al. 2009; Nohynek et al. 2010; Vermeire et al. 2010) and can help to avoid animal use or reduce the number of animals needed for confirmative testing. Inclusion of toxicokinetic data

obtained by modelling and/or combination of in vitro (Blauboer 2002, 2003, 2008) and/or in vivo tests is crucial. Information on exposure can also be incorporated to inform the need for in vivo testing; testing can be avoided if the predicted exposure level is not considered to be significant (i.e. exposure-based waiving), for example, using threshold of toxicological concern (TTC) approaches. A number of integrated testing strategies for repeated dose toxicity testing have been proposed for use under REACH, which could also potentially be applied to cosmetic ingredients (Grindon et al. 2008; ECHA 2008). Different approaches have already been applied, with variable results.

For cosmetics, the idea behind the recent COLIPA research call with respect to repeated dose toxicity testing (see “Current initiatives to develop approaches for repeated dose toxicity”) is to come to an integrated approach by combining the results that will be generated by the funded projects. Indeed, each project is dealing with a specific unsolved problem related to repeated dose toxicity and can be seen as an essential building block in a larger strategy. An example is given for liver by Vanhaecke et al. (submitted for publication to Archives of Toxicology).

In an integrated model, the neurodegenerative properties of acrylamide were studied in differentiated SH-SY5Y human neuroblastoma cells by measuring the number of neurites per cell and total cellular protein content and using a biokinetic model based on acrylamide metabolism in rat (DeJongh et al. 1999b). The hazard assessment was performed on the basis of QSAR, PBBK modelling and new in vitro studies were undertaken. Acute and subchronic toxicity (repeated dose 90-day study) was estimated for rat in vivo and compared to experimentally derived LOELs (lowest observed effect levels) for daily intraperitoneal exposure to acrylamide. The estimated LOELs differed maximally twofold from the respective experimental values and the nonlinear response to acrylamide exposure over time could be simulated correctly. Although the integrated model could predict the toxic dose for an important endpoint after subchronic exposure (altered acoustic startle response), the toxic effects and recovery from toxicity did not mimic completely the situation in vivo. It should also be noted that neurotoxicity is not considered the critical effect driving risk assessment for acrylamide (the critical effect is cancer). However, such a model might be used to roughly predict in vivo toxicity of acrylamide when its concentration in blood serum is known.

Various in vitro and in silico methods were applied to 10 substances for which data from in vivo tests were also available and predictions for a number of effects including acute toxicity, skin and eye irritation and toxicity after repeated dosing were made (Gubbels-van Hal et al. 2005). Acute oral toxicity (LD50) was predicted correctly for 5

out of 10 substances, and skin and eye irritation was predicted for 7 of the 10 substances, but predictions for repeated dose toxicity were correct for only 2 of the 10 substances (Gubbels-van Hal et al. 2005). The evaluation of repeated dose toxicity was based only on one toxicity endpoint, cytotoxicity (in the liver), which may be regarded as a high-dose effect and very crude measure of repeated dose toxicity. In general, the repeated dose toxicity was overestimated by the integrated approach. This study demonstrates the greater difficulty in using alternative methods to predict repeated dose toxicity in comparison with predicting local effects or toxicity after shorter exposures.

Threshold of toxicological concern (TTC) The TTC is a pragmatic risk assessment tool that establishes a human exposure threshold for chemicals without specific toxicity data (Kroes and Kozianowski 2002). The approach is based on the assumption that a conservative estimate of the dose below the toxicity threshold for an untested chemical can be based on the distribution of NOAELs for a large number of tested chemicals (Kroes et al. 2005; Munro et al. 2008). The application of the TTC-concept requires only knowledge of the structure of the chemical and the measured or anticipated exposures. Low exposure levels without appreciable health risks can be identified for many chemicals.

At present, a database containing carcinogenicity data from animal studies (Carcinogen Potency Database, CPDB) containing 730 chemicals and one database containing 613 chemicals based on other toxicological endpoints (Munro database) are available. Both are based on systemic effects after oral exposure. In addition, a database (RepDose) containing toxicity data on 578 industrial chemicals based on oral and inhalation studies is available (Bitsch et al. 2006).

Kroes et al. (2004) describe the work of an Expert Group of ILSI Europe that culminated in the development of a decision tree that is now widely cited as providing the foundation for a tiered TTC approach. This publication describes a step-wise process in which it is first determined whether TTC is an appropriate tool (proteins, heavy metals and polyhalogenated-dibenzodioxins and related compounds have been so far excluded from use with TTC), and then follows a series of questions to determine the appropriate TTC tier. The initial step is the identification of high potency carcinogens that have currently been excluded from the TTC approach (aflatoxin-like, azoxy and N-nitroso compounds). After that, the chemical would be analysed for structural alerts for possible genotoxicity. Those with alerts would be assigned to the lowest TTC tier of 0.15 µg/day.

The TTC approach has primarily been used in a regulatory context for oral exposure to food contact materials, food flavourings and genotoxic impurities

in pharmaceuticals (Kroes et al. 2005). Application of the approach to cosmetic ingredients and impurities requires consideration of whether these values may be applicable to cosmetic exposures by the dermal or inhalation routes. Kroes et al. (2007) reviewed the use of the TTC in safety evaluation of cosmetic ingredients and concluded that the oral TTC values are valid for topical exposures. It was proposed that conservative default adjustment factors can be applied to take account of the relationship between the external topical dose and the internal dose.

In a joint effort with SCHER (Scientific Committee on Health and Environmental Risks) and SCENIHR (Scientific Committee on Emerging and Newly Identified Health Risks), the SCCS (Scientific Committee on Consumer Safety) is currently reviewing the use of the TTC for cosmetic ingredients, and the outcome of this review may have implications for use of the TTC in risk assessment for cosmetics.

More recently, a TTC has been developed for the assessment of inhalation exposure to aerosol ingredients in consumer products. A database of inhalation toxicity information was established and reviewed to derive TTCs for local effects in the respiratory tract and systemic toxicity for Cramer class I and III chemicals. While not all chemicals are suitable for use in the approach, it is proposed that for chemicals with a predictable low potential toxicity and very low levels of exposure, inhalation toxicity testing could be avoided (Carthew et al. 2009).

Industry strategies A number of companies have developed in-house strategies that can be used as practical aids in risk assessment. These strategies generally seek to use all available data on a new ingredient, including information on predicted levels of consumer exposures during product use, as part of a weight of evidence approach to determine whether, and what, new in vivo testing may be required to inform a quantitative risk assessment. Two examples of strategies currently employed that the Group are aware of are described below.

Company 1 Risk-based approaches are employed at Company 1 to ensure animal data are generated only when required to fulfil the requirements of a safety assessment.

This relies on a thorough understanding of human exposure, and this is the starting point for many safety assessments at Company 1, since exposure-based waiving may provide an opportunity to avoid performing new animal tests (Carmichael et al. 2009). Various models are used to predict consumer exposure, making use of data from dietary surveys for food ingredients/contaminants, consumer habits and practise studies for home and personal care ingredients, and even simulated use testing for products that will be applied as a spray.

Knowledge of consumer exposure is used in tandem with various predictive chemistry approaches (e.g. Cramer classification, DEREK) to derive TTCs. TTCs are available for systemic exposure, dermal sensitisation (Safford 2008) and inhalation exposure (Carthew et al. 2009) and are used where appropriate, i.e. where exposure is very low.

For botanical ingredients, history of use is also taken into account using an in-house software tool. This tool takes into account the level of similarity of the proposed ingredient to an historical comparator, plus any evidence for safety or otherwise for the ingredient. This tool is used to model the level of concern associated with the proposed exposure to inform the need for further testing.

Read across to structurally similar chemicals can provide reduction or replacement opportunities. However, in practice, this is rarely used for systemic endpoints, partly due to limited ability to read across to very novel ingredients and partly because a robust evaluation requires a large amount of animal data on structurally similar materials, which is rarely available.

Company 2 The chemical food safety group at Company 2 has developed an effective strategy that integrates a set of in silico models (Fig. 9). Any new molecule entering the system is first tested for mutagenicity using an in-house in silico model (Mazzatorta et al. 2007). In the absence of an alert for mutagenicity/genotoxicity, the calculated exposure is compared to the relevant TTC (Kroes et al. 2004). Exposures lower than the relevant TTC are considered of low safety concern. In case of estimated exposure higher than the relevant TTC, MoEs between the predicted rat LOAEL obtained from TOPKAT, rat LOAEL calculated by an in-house model (Mazzatorta et al. 2008) and the

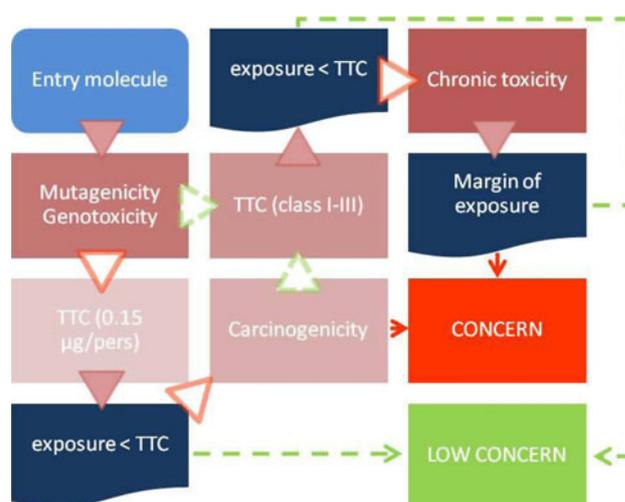


Fig. 9 Integration of QSAR and TTC. Shaded arrows indicate the workflow, open arrows indicate decision outcomes. Green dotted or orange solid arrows correspond to pass or fail in the test, respectively

MRTD model (Maunz and Helma 2008) are calculated. The MoE is calculated by dividing the predicted chronic toxicity values by the estimated exposure. The interpretation of the MoE is performed on a case-by-case basis and takes into consideration interspecies and intraspecies differences. To conclude that there is low concern, MoEs based on rat LOAELs should at least be large enough to account for potential inter- (UF = 10) and intra (UF = 10)-species differences and to allow for the conversion of LOAELs into NOAELs (UF = 3–10). An additional factor would increase the confidence to fully cover the potential error of the models.

In case of alerts for mutagenicity/genotoxicity, the estimated exposure is compared to a lower TTC of 0.15 µg/pers (Kroes et al. 2004). Exposure below this TTC would be considered unlikely to be of any concern, even for compounds with mutagenic properties. Time adjustment may be envisaged in case of established short-duration exposure. Additional development is necessary to handle chemicals with mutagenicity alerts at exposure levels significantly higher than the TTC of 0.15 µg/pers. Models predicting carcinogenicity are currently being evaluated. A chemical with a mutagenicity alert but negative in carcinogenicity predictive models would then enter the chronic toxicity prediction scheme as described previously. A chemical positive in both mutagenicity and carcinogenicity predictions could theoretically be managed through the calculation of a MoE between a predicted carcinogenic potency (e.g. BMDL10) and the estimated exposure. However, no tools are currently available to predict carcinogenic potency in the absence of in vivo data.

Challenges for the development of alternative approaches for quantitative risk assessment of cosmetic ingredients

Quantitative risk assessment

Quantitative human health risk assessment of any chemical substance requires data on the following three elements: hazard identification (potential adverse effects of the substance), dose–response data for each toxic effect (i.e. hazard characterisation) and exposure assessment (the level of exposure preferably in quantitative terms). These data are combined in risk characterisation, i.e. determination of the probability that the toxic effect will occur and the magnitude of the risk. As described earlier, data on specific hazards may be obtained by in vitro studies and other alternative methods for some endpoints. Only in vivo studies provide sufficient information to cover all possible endpoints at present. The dose–response data for the endpoints used in risk assessment are needed for deriving the No observed adverse effect level (NOAEL) of the

substance or another point of departure for risk characterisation. Such data can not yet be obtained in in vitro studies. In risk characterisation, the MOS or MOE is calculated from a comparison of the point of departure (or a health-based guidance value such as a tolerable daily intake) and the level of human exposure.

The SCCS, previously called the Scientific Committee on Consumer Products (SCCP), has recently stated that the evaluation of the systemic risk via repeated dose toxicity testing is a key element in evaluating the safety of new and existing cosmetic ingredients (SCCS 2009). If these data are lacking in a new cosmetic ingredient submission to the SCCS, it is considered not feasible to perform risk assessment of the compound under consideration (SCCS 2009). In addition, it was stated that at present, no alternative methods to replace in vivo repeated dose toxicity testing on experimental animals are available (SCCS 2009).

Understanding the mechanisms of toxicity and modes of action would be a key issue in risk assessment. When they are known for the substance of interest, the relevance of the toxicity to humans may be better evaluated. In vitro studies produce invaluable data on cellular mechanisms for risk assessment. However, in vitro data alone are currently insufficient for quantitative risk assessment.

Combes et al. (2006), Prieto et al. (2006) and recently Boekelheide and Campion (2010) describe how data from in vitro studies could potentially be used in future for risk assessment. The new paradigm focuses on in vitro approaches with human cells, toxicity pathways and high-throughput techniques. Results provide input into a toxicological factors analysis and classification system which can distinguish between adaptive and adverse reactions which are defined at the sub-cellular, cellular and whole organ level. In vitro data on biokinetics will allow establishing physiologically based biokinetic (PBBK) models as a prerequisite for quantitative risk estimation (Blaauboer 2002, 2003). Integration of data on humans can contribute to the weight of evidence (Weed 2005). However, these approaches are at an early stage and have not been evaluated regarding their ability to predict in vivo toxicity. For repeated dose toxicity, much more work is needed for extrapolation of in vitro data towards the in vivo situation (Bessems 2009; see also the overall chapter and that on Toxicokinetics).

Limitations of in vivo studies related to quantitative risk assessment

The goal of toxicity testing is to ensure safety of products and substances in exposed human populations. All experimental non-human and human models have their limitations to reveal and predict human toxicity for safety

evaluation. In considering the development of new approaches for safety testing, it is important to bear these limitations in mind so that they may possibly be overcome.

Regarding the standard repeated dose studies, and in vivo animal studies in general, limitations include

- Mechanisms of toxicity are rarely revealed.
- Species differences in metabolism and toxicokinetics in general, physiological and anatomical differences.
- Genetic polymorphism is not fully covered due to limitations on group sizes in the studies and cross-species differences in polymorphisms.
- Some toxicity targets are poorly evaluated (e.g. cardiotoxicity and neurotoxicity).
- Contribution of age and various disease-linked parameters to toxic response remains unknown.

Some of these limitations are specific to the use of animals as surrogates for humans, while some are due to the study designs employed, for example, it is not feasible (or desirable) to use extremely large numbers of animals in an attempt to detect rare effects. Because the aim of the studies is to evaluate toxicity, the dose range (typically 3 doses of test substance) used is generally higher than human exposures. High doses may cause toxicity which is irrelevant to humans (irrelevant mechanisms and/or endpoints). In addition, extrapolation from high doses to lower exposure level is needed in risk assessment.

Although widely accepted, the procedure to derive MOS values from NOAEL in test animals has not been formally validated for the purposes of predicting human health risks (Blauboer and Andersen 2007).

Only a limited number of surveys from concordance of toxicity in humans and animals have been published from the perspective of repeated dose toxicity studies. Fully comparable data sets are rare or have not been available for evaluation. In a survey on 150 pharmaceutical compounds for which human toxicity was identified during clinical development, the overall concordance rate between effects in humans and those observed in rodent and non-rodent species was 71%, while tests on rodents only predicted 43% of human effects (Olson et al. 2000). The highest incidence of concordance was observed in haematological (80%), gastrointestinal (85%) and cardiovascular effects (80%). For cutaneous effects, the concordance was lowest (about 35%) and for liver toxicity about 55% (Olson et al. 2000). However, it is important to note that this study is limited due to the nature of the drug development process: as it only assessed compounds that reached the market many substances for which toxicity may have been predicted in the animal studies would already have been screened out and not included in the analysis.

In another survey of concordance of hepatotoxic effects between rodents and humans, positive prediction was 60%

for 1,061 pharmaceutical compounds and 46% for another set of 137 compounds (Spanhaak et al. 2008). Between 38 and 51% of liver effects reported in humans were not detected in rodent or non-rodent species. At least some of these liver effects may have been idiosyncratic effects that also would not have been detected in pre-marketing trials because of their rarity.

Altogether, toxicity testing in animal models cannot reveal all potential toxicity in humans. However, even studies in human subjects, such as in phase I or phase II clinical trials, are unable to reveal all potential adverse effects in humans because of the limited numbers involved.

It is important to bear in mind when developing alternatives that the ultimate goal is prediction of health effects in humans, and there is thus a need to change the mindset from trying to mimic animal data and one-to-one replacement for each target organ.

Limitations of in vitro studies with specific cell types

Endpoints relevant for regulatory decisions are usually based on NOAELs¹⁴ from animal studies. Despite the many parameters assessed in these studies, pathologic changes are currently the most relevant for deriving NOAELs. The pathological changes usually are widely varying and chemical specific, and both target organ and cell-specific effects within an organ are frequent. Pathological changes often involve complex interactions between different cell types in a specific organ and the pathological response may be mediated or influenced by mediators released by other tissues such as the immune, the inflammatory or the endocrine system.

In contrast, toxicity assessment in cultured cells usually is based on observations of toxicity or functional effects in a single-cell type, of which the biochemistry and gene expression may or may not be similar to those of the target cell in vivo. Responses to chemical challenges may also differ. Since many of the mechanisms leading to toxicity are only poorly defined at present, it is highly questionable whether toxicity testing in vitro using specific cellular systems will be predictive for the complex integrated pathological responses in vivo.

Furthermore, there are currently no readily available/reliable ways to quantitatively extrapolate dose–response data from concentrations of the test substance in in vitro experimental systems to toxic dose/exposure level in the whole body. Physiologically based pharmacokinetic (PBPK) and physiologically based biokinetic (PBBK) models may be applicable to predict concentrations of test

¹⁴ Whilst alternatives such as the BMDL10 have been proposed, they offer only a modest refinement in the risk assessment and most of the issues relevant to the NOAEL also apply to the BMDL10.

substance and their metabolites in tissues, although these approaches are data-intensive.

Overall, appropriate information on the dose response of adverse effects, identification of thresholds and NO(A)ELs that are essential for risk characterisation cannot be obtained at present from *in vitro* studies (Greim et al. 2006). In order to improve the predictivity of *in vitro* systems, major efforts to understand mechanisms of toxicity on a tissue/organ level are required. Only then can relevant biochemical factors contributing to toxicity be identified and integrated in the *in vitro* test systems. Some optimistic views already exist about the possibilities of using, in future, non-animal data as input for quantitative risks assessment in particular when well-defined threshold dose levels can be defined *in vitro* (Combes et al. 2006).

From the perspective of their usefulness for evaluation of repeated dose toxicity, *in vitro* studies are at present most suitable for:

- elucidating the cellular mechanisms of toxicity
- providing additional data on potential to cause specific toxic effects (hazard identification)
- relative ranking of hazard potency

The use of human-based models may also be an advantage for predicting human effects.

Importance of understanding of mode of action and toxicity pathways in the development of alternative approaches

The development of the concept of mode of action (MOA), initially in chemical carcinogenesis and subsequently extended to systemic toxicity in general, has had a marked impact on many aspects of toxicology, not least in providing a clear focus and rationale for the development of alternatives to toxicity testing in experimental animals. A MOA comprises a series of key events, i.e. effects that are necessary, though individually probably not sufficient, to cause toxicity, observable and quantifiable. A mode of action usually starts with the interaction of an agent with a cell, through functional and anatomical changes, resulting in an adverse effect on health (Boobis et al. 2009a). An example of a key event would be the biotransformation of methanol to formaldehyde. While this is a necessary event in the ocular toxicity of methanol, on its own it is not sufficient. The ocular toxicity (blindness) caused by methanol requires subsequent conversion of formaldehyde to formate, mitochondrial dysregulation and retinal cell loss—all key events. A key event may be either kinetic or dynamic. A MOA, as defined here, can be distinguished from a mechanism of action, which comprises a detailed molecular description of a key event(s) in the induction of an adverse health effect.

By focusing on MOA, it should be possible to develop alternative testing methods, in which key events are assessed and quantified. The advantage here is that the human relevance of a MOA to humans, and of the specific key events, is known in advance. Application of the International Programme on Chemical Safety (IPCS) human relevance framework provides a systematic and transparent means of establishing a MOA in experimental animals and for assessing its human relevance (Boobis et al. 2006, 2008). An important advantage of the MOA approach is that once the human relevance for a MOA has been established for one substance, the implications are applicable to other substances that share the mode of action. Key events can be assessed in simpler systems than are necessary for apical endpoints. For example, in the case above of methanol ocular toxicity, the key event of conversion to the intermediate metabolite formaldehyde can be assessed in a less complex system than the whole organism. Such studies need to be allied to appropriate biokinetic considerations to enable *in vivo* extrapolation. The test system used can be designed to be human relevant in advance, based on knowledge of the mode of action and its key events. By assessing the important key events in a MOA, it may be possible to reach conclusions on the potential consequence for human health without the necessity of performing whole-animal studies. For example, such an approach has been suggested for the assessment of carcinogenic potential, albeit currently using a refinement of animal tests rather than their replacement (Boobis et al. 2009b; Cohen 2004). Nevertheless, the general principle is such that it might be possible to develop non-animal approaches.

The integration of omics studies allied to pathway analysis will improve the predictive capacity of assays for key events. Such next generation omics studies will have established the biological relevance of the pathways identified and will have determined their quantitative contribution to phenotypic changes. Thus, omics will add evidence-based value to the alternative systems being developed. An example of this can be seen in the assessment of genotoxicity, where the pathways affected at the transcriptomic level reflect relevant biological process, distinct from those influenced by non-genotoxic compounds (van Delft et al. 2005).

Identification of key events involved in human-relevant modes of action provides an efficient and effective means of developing biomarkers of effect applicable to studies in humans at exposures below those which are overtly toxic. The biomarker may comprise a key event itself, for example circulating levels of a toxic metabolite, or a suitable surrogate of a key event, e.g. activation of nuclear receptors. While it is not often possible to quantify such activation directly *in vivo* in humans, a gene product

dependent on this activation may provide an adequate surrogate. As an illustration, activation of the aryl hydrocarbon receptor (AhR) cannot be measured directly *in vivo*, whereas products of AhR-regulated genes such as CYP1A2, as assessed by the clearance of a specific substrate such as caffeine, can readily be quantified.

Current initiatives to develop alternative approaches for repeated dose toxicity

There is a growing view that recent advances in science and technology in medicine and the biosciences could be harnessed to develop novel, innovative ways to assure safety without using animals, while also providing a better prediction of likely effects in humans. For example, the US National Research Council (NRC) recently set out a long-term vision for the future of toxicity testing and risk assessment in the twenty-first century, where safety testing would be based on the identification of key biological pathways that, if perturbed sufficiently, would result in harmful effects (NRC 2007). This is somewhat analogous to the key events/mode of action strategy discussed earlier. Chemicals would be evaluated for their potential to produce changes in these key ‘toxicity pathways’:

“Advances in toxicogenomics, bioinformatics, systems biology, epigenetics and computational toxicology could transform toxicity testing from a system based on whole-animal testing to one founded primarily on *in vitro* methods that evaluate changes in biological processes using cells, cell lines or cellular components, preferably of human origin”.

In the USA, the National Toxicology Program (NTP), Environmental Protection Agency (EPA) and National Institutes of Health Chemical Genomics Centre (NCGC) have formed a collaborative research programme to help realise the NRC’s vision in practice. The Tox21 programme seeks to identify cellular responses to chemical exposures that are expected to result in adverse effects and develop high-throughput screening tools that can be used to predict toxicity *in vivo* (Schmidt 2009). Initially, the tools developed will be used to support the selection of previously untested chemicals that should be prioritised for animal testing. In addition, the Human Toxicology Project Consortium is seeking to engage a wide range of stakeholders in academia, industry, regulatory and non-governmental organisations to work together in leading research to facilitate the global implementation of the recommendations of the NRC report. The Consortium is also seeking to develop methods of integrating and interpreting data derived with alternative assays so they can be used for quantitative risk assessment.

In Europe, the European Commission recently had a joint research call, co-funded by COLIPA, to help develop technologies and approaches for non-animals methods of assessing repeated dose toxicity. Applications were invited in the following areas:

- Optimisation of current methodologies and development of novel methods to achieve functional differentiation of human-based target cells *in vitro*.
- Exploitation of organ-simulating cellular devices as alternatives for long-term toxicity testing.
- Establishment of endpoints and intermediate markers in human-based target cells, with relevance for repeated dose systemic toxicity testing.
- Computational modelling and estimation techniques.
- Systems biology for the development of predictive causal computer models.
- Integrated data analysis and servicing.

Up to EUR 50 million will be invested in these projects, which are expected to start in 2011. This research and the other activities discussed above are intended to contribute to the development of alternative methods that can form building blocks in the integrated approach needed for quantitative risk assessment. However, it is not currently possible to predict when full replacement of animals might be achieved; the NRC’s vision anticipates that it will take several decades.

Conclusions

- No complete replacement for repeat-dose toxicity will be available by 2013. Development of alternatives in this area is extremely challenging, and it is not possible to predict when full replacements are likely to be available.
- *In vitro* tests can enable hazard identification for certain specific endpoints and data on mechanisms of action and potential toxicity.
- *In silico* models are already extensively used by the pharmaceutical industry to guide and prioritise new chemicals/drugs discovery and development.
- QSARs, read-across from analogue chemicals, exposure-based approaches (including TTC for trace amounts) and *in vitro* predictive toxicology screening are being increasingly used in intelligent testing strategies in a weight of evidence approach to exclude possible hazards to health for a range of substances or cosmetic applications. With further development, such approaches could make an even greater contribution to avoiding animal testing by 2013, even if full replacement cannot be expected.

- Key gaps include the following:
 - In vitro assays to detect apical endpoints in target organs, which have been the major focus for development of alternative methods to date, are unlikely to be sufficient for complete replacement. For quantitative risk assessment using alternative methods, there is a need for better and more scientific knowledge on exposure, toxicokinetics and dose response, mechanisms of toxicity and extrapolation between exposure routes. Better understanding of MOA and key events associated with repeated dose toxicity endpoints would support development of alternative approaches.
 - The main bottleneck for developing predictive QSAR models is the availability of sufficiently large high-quality training datasets. These models and their applicability domain could be greatly improved by promoting the publication of confidential data.
 - One of the major challenges is to reproduce integrated responses. There is a need to develop approaches/strategies for combining and interpreting data on multiple targets/endpoints, obtained from a variety of alternative methods, so they can be used in human risk assessment.
 - Methods for dose response extrapolation from in vitro to in vivo are therefore needed. There is also a need to improve and expedite approaches for obtaining relevant data and for development of informative models.
 - It is important to bear in mind that the ultimate goal is prediction of health effects in humans, and there is thus a need to change the mindset from trying to mimic animal data and one-to-one replacement for each target organ towards a human-relevant integrated approach combining all possible knowledge and integrating, whenever possible, quantitative data with evidence-based findings.
- apparatus and, non-genotoxic carcinogens that exert their carcinogenic effects through other mechanisms.
- The 2-year cancer bioassay in rodents is widely regarded as the gold standard to evaluate cancer hazard and potency; however, this test is rarely done on cosmetic ingredients. A combination of shorter-term in vitro and in vivo studies has been used including in vitro and in vivo genotoxicity assays, to assess genotoxic potential and repeat-dose (typically 90-day) toxicity studies to assess the risk of non-genotoxic chemicals.
- It is clear that the animal testing bans under the 7th Amendment to the Cosmetics Directive (EU 2003) will have a profound impact on the ability to evaluate and conduct a quantitative risk assessment for potential carcinogenicity of new cosmetic ingredients. This impact is not only due to the ban on the cancer bioassay itself, but mainly to the ban on in vivo genotoxicity testing, any repeat-dose toxicity testing, and other tests such as in vivo toxicokinetics studies and in vivo mechanistic assays which are currently used to aid safety assessment.
- This report is a critical evaluation of the available non-animal test methods and their ability to generate information that could be used to inform on cancer hazard identification.
- Although several in vitro short-term tests at different stages of development and acceptance are available, at the current status these will not be sufficient to fully replace the animal tests needed to confirm the safety of cosmetic ingredients. Furthermore, they are focused on hazard evaluation only and cannot currently be used to support a risk assessment.
- However, for some chemical classes, the available non-animal methods might be sufficient to rule out carcinogenic potential in a weight of evidence approach.
- Taking into consideration the present state of the art of the non-animal methods, the experts were unable to suggest a timeline for full replacement of animal tests currently needed to fully evaluate carcinogenic risks of chemicals. Full replacement is expected to extend past 2013.

Carcinogenicity

Executive summary

- Carcinogenesis is a complex long-term multifactorial process and consists of a sequence of stages.
- Carcinogens have conventionally been divided into two categories according to their presumed mode of action: genotoxic carcinogens that affect the integrity of the genome by interacting with DNA and/or the cellular

General considerations

Introduction

Substances are defined as carcinogenic if after inhalation, ingestion, dermal application or injection they induce (malignant) tumours, increase their incidence or malignancy or shorten the time of tumour occurrence. It is generally accepted that carcinogenesis is a multistep/multihit process from the transition of normal cells into cancer cells via a sequence of stages and complex biological

interactions, strongly influenced by factors such as genetics, age, diet, environment, hormonal balance, etc.

Since the induction of cancer involves genetic alterations which can be induced directly or indirectly, carcinogens have conventionally been divided into two categories according to their presumed mode of action: genotoxic carcinogens and non-genotoxic carcinogens.¹⁵ Genotoxic carcinogens have the ability to interact with DNA and/or the cellular apparatus (such as the spindle apparatus and topoisomerase enzymes) and thereby affect the integrity of the genome, whereas non-genotoxic carcinogens exert their carcinogenic effects through other mechanisms that do not involve direct alterations in DNA.

The 2-year cancer bioassay in rodents is widely regarded as the gold standard to evaluate cancer hazard and potency, although it is generally known that this test has its limitations to predict human cancer risk (Knight et al. 2005, 2006). However, this test is rarely done on cosmetic ingredients. Rather, a combination of shorter-term *in vitro* and *in vivo* studies has been used including *in vitro* and *in vivo* genotoxicity tests to assess genotoxic potential and repeat-dose (typically 90-day) toxicity studies to assess non-genotoxic potential.

It is clear that the animal testing bans under the 7th amendment of the Cosmetics Directive (EU 2003) will have a profound impact on the ability to evaluate and conduct a risk assessment for potential carcinogenicity of new cosmetic ingredients. This impact is not only due to the ban on the cancer bioassay itself, but also to that on *in vivo* genotoxicity testing, any repeat-dose toxicity testing, and other tests such as toxicokinetics studies and *in vivo* mechanistic assays that currently can be used to aid safety assessment.

The challenge will be to find/develop alternative tests for both genotoxic and non-genotoxic carcinogens. The complexity of the carcinogenicity process makes it difficult to develop *in vitro* alternative test models that mimic the full process, especially for non-genotoxic chemicals. The challenge in developing *in vitro* alternatives is also heightened because of the complexity of the number of potential target organs. Some key events of the carcinogenesis process can be investigated *in vitro*. However, it is expected that an integrated approach involving multiple *in vitro* models will be needed, but a better understanding of the entire process is needed before this will be possible. Scientific research is ongoing to try to achieve this goal.

¹⁵ It is noted that not all genotoxic events lead to mutagenicity, and that some prefer the terminology “mutagenic mode of action.” However, genotoxicity assays are still commonly used to distinguish those chemicals with the potential to directly affect the integrity of DNA from those that don’t, so for the sake of simplicity, the text throughout refers to genotoxic versus non-genotoxic carcinogens.

Information requirements for the carcinogenic safety assessment of cosmetics ingredients until March 2009 (ref. SCCP notes of guidance)

The EU Scientific Committee on Consumer Products (SCCP) issued the 6th revision of the “Notes of Guidance for the Testing of Cosmetic Ingredients and their Safety Evaluation” in 2006 (SCCP 2006). This Guidance document lists the general toxicological requirements for the submission of a cosmetic ingredient dossier to the SCCP as follows: acute toxicity, irritation and corrosivity, skin sensitisation, dermal/percutaneous absorption, repeated dose toxicity and genotoxicity. These are considered the minimal base set requirements. However, when considerable oral intake is expected or when the data on dermal absorption “indicate a considerable penetration of the ingredients through the skin (taking into account the toxicological profile of the substance and its chemical structure), additional data/information on carcinogenicity, reproductive toxicity and toxicokinetics may become necessary, as well as specific additional genotoxicity data”. It is noted that the SCCP Notes of Guidance does not define what is meant by “considerable oral intake” or “considerable penetration of the ingredients through the skin”. Tools such as the Threshold of Toxicologic Concern (TTC) may be helpful in determining which exposures warrant toxicological evaluation.

Historically, the strategy for addressing the carcinogenicity endpoint for cosmetic ingredients has been threefold:

1. First, compounds are evaluated for genotoxicity. The first step in this evaluation was a battery of *in vitro* genotoxicity tests. A positive finding in an *in vitro* test (e.g. a chromosome aberration study, OECD 473) could then be followed-up with an *in vivo* study (e.g. mouse micronucleus, OECD 474), which is deemed to have more relevance to human exposures. Positive *in vivo* tests for their part could trigger carcinogenicity testing. In general, compounds that have been shown to have genotoxic potential *in vivo* are not used in the formulation of cosmetics, and materials testing positive in these tests have rarely been pursued further as this might require the conduct of a carcinogenicity bioassay (OECD 452) or a combined chronic toxicity/carcinogenicity test (OECD 453). These studies take several years to run (the *in-life* portion alone lasts 24 months) and cost around one million Euro. For these reasons and given the potential for genotoxic compounds to be positive in a rodent bioassay, new cosmetic ingredients are almost never tested in a carcinogenicity bioassay.
2. For those chemicals shown to lack genotoxicity potential, it is generally assumed that there is a

threshold and that the carcinogenic risk can be avoided based on data from repeat-dose toxicity studies. Prior to the formation of tumours (generally seen only after long-term exposures), non-genotoxic carcinogens cause changes in normal physiological function, and these adverse effects, if relevant to the exposure, would be determined in a sub-chronic study. Accordingly, the risk assessment generally involves the identification of a NOAEL from an appropriate repeat-dose toxicity study (e.g. 90-day study) and the application of appropriate safety factors. The methods used in such quantitative risk assessments are regarded as being sufficiently conservative such that even if the chemical was later shown to be a non-genotoxic carcinogen, the exposure would be so low that there would be no risk to consumers. This is consistent with the risk assessment practices of virtually all regulatory bodies, including those inside and outside of Europe.

3. In addition to repeat-dose toxicity studies, other *in vivo* studies are sometimes used to better understand the human relevance of findings in rodents (e.g. related to toxicokinetic handling or species-specific effects) or the mechanism and associated dose–response for a chemical.

Implications for carcinogenic safety assessment after the 7th Amendment

The 7th Amendment to the Cosmetics Directive (EU 2003) banned the conduct of all *in vivo* tests for cosmetic ingredients in EU starting March 2009. However, a full marketing ban of ingredients tested for repeat-dose toxicity (including carcinogenicity and sensitisation), reproductive toxicity and toxicokinetics, which tests are still allowed for ingredients outside the EU, is foreseen in March 2013. The consequences of these bans for carcinogenicity assessment are that (1) for genotoxic substances, no *in vivo* genotoxicity tests as follow-up of *in vitro* positive tests are allowed and (2) risk from non-genotoxic carcinogens cannot be sufficiently evaluated since the repeated dose toxicity (and carcinogenicity) test is no longer allowed. Since both modes of action are important and need to be covered, for both alternative methods and approaches should be considered.

Assessment of genotoxic carcinogens

Until the 7th Amendment, both *in vitro* and *in vivo* tests played an important role in the recognition of potential carcinogenicity in cosmetic ingredients. The *in vivo* tests were used to clarify whether positive results from *in vitro* tests were relevant under conditions of *in vivo* exposure. A

number of well-established and regulatory accepted *in vitro* tests are in place, but a caveat to the use of these tests is the relatively low specificity and high rate of misleading positive results (i.e. the results are not indicative of an increased cancer risk associated with DNA reactivity, as generally assumed from these tests). Kirkland et al. (2005, 2006) evaluated the predictivity of standard *in vitro* tests for rodent carcinogens. The combination of three *in vitro* genotoxicity tests as required by the SCCS increases the sensitivity of the test battery (up to 90%), but the specificity (ability to identify non-carcinogens) decreased drastically (down to below 25%). The low level of specificity means that unacceptably high numbers of positives are generated. However, before March 2009, the positives could be evaluated and often overruled with *in vivo* genotoxicity tests. Thus, the ban of *in vivo* tests will have a negative impact on the development of new cosmetic ingredients. This is clearly demonstrated by the evaluation of 26 hair dyes by the SCCP. Nineteen hair dyes had to be further assessed due to a clastogenic effect found in the *in vitro* systems. For these compounds, 37 *in vivo* genotoxicity tests were submitted and 35 turned out to be negative and 2 (comet assays) were at the time considered equivocal. These data indicate that the performance of at least 26 *in vivo* tests was deemed necessary for the appropriate characterisation of the genotoxic potential. Without the performance of *in vivo* tests, at least 17 of these hair dyes might have been abandoned without full scientific justification (Speit 2009). Identical results were obtained from a comprehensive survey of all SCC(NF)P/SCCS opinions issued between 2000 and 2009. It appeared that 97 compounds (of 150 tested ones) would have been lost without the presence of the *in vivo* data (Pauwels and Rogiers 2009). However, we acknowledge that some *in vivo* genotoxicity tests (e.g. UDS test) may suffer from limited sensitivity (Benigni et al. 2010; Kirkland and Speit 2008).

Based on the weak performance of existing *in vitro* tests, the development of new *in vitro* tests with better predictivity for cosmetic ingredients is in focus. An ECVAM workshop on “how to reduce false positive results when undertaking *in vitro* genotoxicity testing” was held in 2006 and identified, among others, the following factors as being important for the improvement of the current tests (Kirkland et al. 2007): (1) the identification of the chemicals to be used in the evaluation of modified or new tests (Kirkland et al. 2008), (2) the choice of the cell types with higher relevance (e.g. human origin, p53-proficient), (3) the current measures in cytotoxicity in the standard genotoxicity tests and (4) the current maximum concentration in the standard genotoxicity tests.

Although inside the EU *in vivo* genotoxicity tests are not allowed anymore, until 2013 they can still be carried out outside the EU under some circumstances. For example,

outside the EU, the micronucleus test and the comet assay can still be performed if included in a repeated dose study (Pfuhler et al. 2009). The *in vivo* comet assay gained growing scientific acceptance over the last years, whereas the inclusion of the micronucleus test is a long and well-established concept and is already represented in the current OECD guideline (OECD 474). While the integration of a micronucleus test into repeated dose toxicity studies can be accomplished without the addition of a positive control, this may be problematic for other tests. For example, the problem with the comet assay may be that this currently requires the inclusion of an endpoint-specific positive control group (Pfuhler et al. 2009). The use of positive control reference slides, however, could be an alternative to control for technical variations during study performance. In case the repeated dose studies which include a measure of genotoxicity but not an endpoint-specific positive control will not be accepted by the regulatory bodies and thus cannot be used, the only way to further evaluate positive findings from *in vitro* genotoxicity studies in order to clarify the possible carcinogenic potential of a compound is the performance of a carcinogenicity study. This underscores the value of having a control group scientifically to answer questions that could only otherwise be answered by a bioassay.

The ability to investigate the relevance of positive *in vitro* genotoxicity results for prediction of carcinogenicity in humans without the use of animals is a significant scientific and technical challenge. In addition to an improvement of the existing *in vitro* genotoxicity tests, a range of scientifically accepted tools should be available to allow appropriate experiments for *in vitro* follow-up testing.

Assessment of non-genotoxic carcinogens

Although it is generally accepted that major carcinogenic risk is related to genotoxic compounds which can well be detected by *in vitro* methods, the potential risk related to non-genotoxic compounds must also be evaluated. Despite the fact that some of the major mechanisms behind non-genotoxic carcinogenicity are known, multiple unknown mechanisms of action and the insufficient knowledge of the cellular and molecular events have not yet allowed for the implementation of a battery of *in vitro* tests that could predict and/or explain their carcinogenic potential to man.

The mechanisms by which non-genotoxic carcinogens cause tumours are in most cases related to tissue- and species-specific disturbances in normal physiological control, gene expression patterns implicated in cellular proliferation, survival and differentiation (Widschwendter and Jones 2002; Baylin and Ohm 2006; Esteller 2007). Numerous examples exist where the mechanism is animal

species specific, and thus these effects found in animals, consequently, are not predictive for humans (Shanks et al. 2009).

The mechanisms behind non-genotoxic carcinogenicity can be manifold, many of which are still not completely understood. Typically modes of action are related to carcinogenesis phases of promotion and progression, but participation in initiation phase is also proposed (Hattis et al. 2008). The induction of tissue-specific toxicity (cytotoxicity) resulting in inflammation and regenerative hyperplasia belongs to those of the well known. Chronic inflammation has shown to be associated with increased incidence of cancers (Loeb and Harris 2008). As a result of cell death by cytotoxic agents, persistent regenerative growth may occur with increasing probability for spontaneous mutations (Ames and Gold 1991) which may lead to accumulation and proliferation of mutated cells giving rise to pre-neoplastic foci and, ultimately, to tumours via further clonal expansion. Induction of immunosuppression by chemicals is regarded another significant non-genotoxic mechanism of cancer. The results from immunosuppressant drugs like cyclosporine A have shown that they can elicit direct cellular effects that can lead to promotion of cancer, independent of immune reactivity (Hernández et al. 2009). Oxidative stress in cells also results in non-genotoxic carcinogenesis as it is shown that cancer cells commonly have increased levels of reactive oxygen species (ROS) and that ROS can induce cell malignant transformation (López-Lázaro 2010; Klaunig et al. 2010). Oxidative stress has been suggested to have some involvement in the mode of carcinogenic activity of peroxisome proliferators in rodent livers (Doull et al. 1999; Hernández et al. 2009).

Many non-genotoxic carcinogens act via binding to receptors such as aryl hydrocarbon, nuclear and peroxisome-proliferator receptors (Hattis et al. 2008), thus affecting proliferation, apoptosis and intercellular communication. Relevant roles are also granted for tyrosine kinase (TK), ion channel-coupled and G-protein-coupled receptors (Lima and van der Laan 2000). Many endocrine modifiers act through hormonal-mediated processes by binding to receptors such as the oestrogen, the progesterone, the aryl hydrocarbon or the thyroid hormone receptors and induce cell proliferation at their target organs (Lima and van der Laan 2000). Chemical substances may also cause tumours by affecting regulation of gene expression and genomic stability through hyper- or hypomethylation of DNA, histone modifications and nucleosomal remodelling (Lo and Sukumar 2008; Sadikovic et al. 2008).

Different research methods, including *in vitro* methods using several cell types, are available to study a number of these potential mechanisms. For example, tests are available to measure oxidative stress (Klaunig et al. 2010) or to measure the inhibition of gap junction intercellular

communication (GJIC) (Klaunig and Shi 2009), both of which have been associated with a number of non-genotoxic carcinogens. However, these methods cannot currently be used to reliably predict carcinogenic potential, but rather are focused on better understanding the mechanism for effects elicited by a chemical.

Up to now the safety of non-genotoxic compounds in man has mainly been concluded from repeat-dose toxicity tests (mainly 90 day study), toxicokinetics, 2-year carcinogenesis bioassay, if available, and by using the TTC principle. The prerequisite for the adequate use of TTC is that information on toxicokinetics (i.e. systemic exposure) is available. Opposite to genotoxic carcinogens for non-genotoxic carcinogens, the threshold principle is commonly used for risk assessment. At this moment, no in vitro test battery is recommended to test non-genotoxic carcinogenic potential of chemical substances. To avoid animal-specific and biased results, an in vitro testing battery based on human cell or tissue models with relevant biomarkers is seen as the most optimal way to replace animal tests in non-genotoxic carcinogenic risk assessment. It is expected that there will be significant synergies between work to develop replacement tests for repeat-dose toxicity studies with tests to predict non-genotoxic carcinogens and quantitative response thresholds.

Inventory of alternative methods currently available

Currently available non-testing and in vitro methods are described below and are summarised in Table 10.

Non-testing methods

Non-testing methods include (quantitative) structure–activity relationships ([Q]SARs) and the formation of chemical categories to facilitate the application of read-across between similar chemicals. Non-testing methods are based on the assumption that the information of a certain compound can be extracted from the analysis of the effects of similar compounds. Such methods are generally computer-based (in silico) approaches.

(Q)SAR models link toxicity to continuous parameters (molecular descriptors) associated with the chemical structure. In case the relationship is simply qualitative, the expression SAR is used. The term (Q)SAR is an umbrella term covering both cases.

(Q)SARs are often incorporated, possibly in conjunction with databases, into expert systems. An expert system is any formalised system that is often, but not necessarily, computer-based and that can be used to make predictions on the basis of prior information (Dearden et al. 1997). Expert systems (and their implementation in software tools) are based on three main modelling approaches

referred to rule-based, statistically based or hybrid methods. Rule-based methods codify the human rules which identified certain chemical fragments responsible for the effect. Statistical models are built by using data mining methods to extract the information from a set of compounds. It is possible to combine both approaches within a hybrid model.

Quantitative structure–activity relationship (QSAR)

Short description, scientific relevance and purpose: To date, hundreds of QSAR models have been published in the literature for predicting genotoxicity and carcinogenicity. The application of the gene mutation test in bacteria (Ames test) to large numbers of chemicals has shown that this test has a high positive predictivity for chemical carcinogens (around 80%; Benigni et al. 2010). Consequently, the most commonly modelled test for genotoxicity has been the Ames test. Most models are qualitative (SARs), i.e. coarse-grain classifiers that predict a chemical compound as genotoxic or carcinogenic or not. Relatively few models are quantitative (QSARs) which provide a more precise means of assessing genotoxicity and carcinogenicity, mainly for congeneric sets of chemicals.

QSARs for non-genotoxic carcinogenicity are still in an early stage of development. A number of structural alerts and characteristics of several types of non-genotoxic carcinogens have been summarised (Woo and Lai 2003). Relatively few models are available for identifying non-genotoxic carcinogens or for predicting carcinogenic potency (Toropov et al. 2009).

There exist several commercial as well as free available expert systems for predicting genotoxicity and carcinogenicity (Benfenati et al. 2009; Serafimova et al. 2010). Freely available models in the public domain include CAESAR, Toxtree, OncoLogic and LAZAR. Commercial models requiring license fees include MultiCase, TOP-KAT, HazardExpert, DEREK and ToxBboxes.

Rule-based systems contain “if-then-else” rules that combine toxicological knowledge, expert judgment and fuzzy logic. Commonly used software tools based on this approach include OncoLogic (Woo et al. 1995), Derek (Sanderson and Earnshaw 1991; Ridings et al. 1996) and HazardExpert (Smithing and Darvas 1992). Derek and HazardExpert can be used in conjunction with their sister programs Meteor and MetabolExpert to predict the genotoxicity and carcinogenicity potential of metabolites as well as parent compounds. In addition to these commercial tools, models included in Toxtree and the OECD Toolbox (OECD 2010d) are rule based.

Statistically based systems use a variety of statistical, rule induction, artificial intelligence and pattern recognition techniques to build models from non-congeneric databases. Statistically based systems are included in the

Table 10 Summary of identified alternative non-animal methods for carcinogenicity

Current endpoints addressed in animal test	Alternative tests available	Part of mechanism covered	Area(s) of application	Status (R&D, optimisation, prevalidation, validation, regulatory acceptance)	Comments	Estimated time until entry into pre-validation
Genotoxicity and carcinogenicity	QSAR and expert systems	Mainly genotoxicity and carcinogenicity based on DNA reactivity. Mostly classification models	Mainly screening, but also provide a means of filling data gaps in hazard assessment	Models and software tools implementing them are subject to ongoing development	These are in silico tools that generate predictions based on chemical structure; they are not experimental test methods	N/A
Genotoxicity and carcinogenicity	Category formation and read-across approach	Depends on choice of category members	Grouping and read-across provides a means of filling data gaps in both hazard and risk assessment	Various software tools are available to carry out grouping and read-across	These are in silico tools that support grouping and read-across	N/A
Genotoxic and non-genotoxic carcinogenicity	TTC approach	TTC is a statistically based approach to establish a conservative default risk value based on worst-case assumptions about the chemical in the absence of data. It is not an assay that is intended to replace current testing strategies	Screening; conservative risk assessment tool based on distribution of risk values (one distribution for genotoxic compounds and a separate distribution for non-genotoxic compounds)	TTC has regulatory acceptance as a risk assessment tool in the US for food packaging materials and in the US and Europe to set acceptable exposure limits for genotoxic impurities in drugs. It has not yet been granted regulatory acceptance for use in cosmetics	Although TTC has the potential to be a useful risk assessment tool for very low-level exposures, it is not expected to be useful for cosmetic ingredients used at higher levels or in product types associated with higher consumer exposures (e.g. body lotion)	N/A
Genotoxicity	Bacterial reverse mutation assay in <i>Salmonella typhimurium</i> and <i>Escherichia coli</i>	Gene mutations (point mutations, base pair substitutions and frameshift mutations)	Screening Hazard identification	OECD TG 471		Guideline available
Genotoxicity	In vitro gene mutation assay in mammalian cells	Gene mutations (point mutations, base pair substitutions, frameshift mutations) Structural and numerical chromosome damage (Mouse Lymphoma L5178Y cells)	Screening Hazard identification	OECD TG 476	Problems related to low specificity of the tests (exception of <i>hprt</i>)	Guideline available
Genotoxicity	In vitro micronucleus assay in mammalian cells	Structural and numerical chromosome damage in mammalian cells (i.e. clastogenicity and aneuploidy)	Screening Hazard identification	OECD TG 487	Problems related to low specificity of the test	Guideline available

Table 10 continued

Current endpoints addressed in animal test	Alternative tests available	Part of mechanism covered	Area(s) of application	Status (R&D, optimisation, prevalidation, validation, regulatory acceptance)	Comments	Estimated time until entry into pre-validation
Genotoxicity	In vitro chromosome aberration assay in mammalian cells	Structural and numerical chromosome damage in mammalian cells (i.e. clastogenicity and polyploidy)	Screening Hazard identification	OECD TG 473	Problems related to low specificity of the test	Guideline available
Genotoxicity	Comet assay in human 3D skin model	DNA damage	Hazard identification Clarification of in vitro genotoxic positives	Optimisation	Additional work is ongoing with further human 3D skin models	1–3 years
Genotoxicity	Micronucleus assay in human 3D skin model	Structural and numerical chromosome damage in mammalian cells (i.e. clastogenicity and aneuploidy)	Hazard identification Clarification of in vitro genotoxic positives	Prevalidation ongoing for one of the epidermis human models	Additional work is ongoing with further human 3D skin models	0–3 years
Genotoxicity	GreenScreen HC	<i>p53</i> -dependent upregulation of GADD45a expression in response to DNA damage	Screening	Optimisation		1–3 years
Genotoxicity	Hens egg test for micronucleus induction (HET-MN)	Structural and numerical chromosome damage in mammalian cells (i.e. clastogenicity and aneuploidy)	Screening Hazard identification	Optimisation		0–4 years
Genotoxic and non-genotoxic carcinogenicity	Cell transformation assays (CTA)	Transformation	Screening Clarification of in vitro genotoxic positives Hazard identification Identifying promoters Chemopreventive activity Mechanistic studies	Prevalidation completed for SHE and Balb/c 3T3 validation ongoing for Bhas 42	Standard protocol developed for SHE, to be considered for drafting the OECD TG	0 year
Genotoxic and non-genotoxic carcinogenicity	In vitro toxicogenomics	Expression changes of genes and gene sets (biological pathways)	Screening Clarification of in vitro genotoxic positives Hazard identification Mechanistic studies	R&D Optimisation		0–5 years

N/A Not applicable

This table estimates the time needed to enter prevalidation, according to ECVAM's established criteria, assuming optimal conditions. It does not indicate the time needed to achieve full replacement of the animal test, nor does it include the time needed to achieve regulatory acceptance. "Optimal conditions" means that all necessary resources, for example, technical, financial, human and coordination, are met at all times in the process, and that the studies undertaken have successful outcomes

commercial tools MultiCASE (Klopman and Rosenkranz 1994), TOPKAT (Enslein et al. 1994) and the publicly available LAZAR (Helma 2006) and CAESAR (Ferrari and Gini 2010; Fjodorova et al. 2010) models. In addition, many models published in the literature and not implemented in software are statistically based.

Hybrid models are based on a combination of knowledge-based rules and statistically derived models. These are based on the general idea that within the structural space of a single structural alert (considered to represent a single interaction mechanism), statistically derived models can quantitatively predict the variation in the reactivity of the alert conditioned by the rest of the molecular structure. Examples of the hybrid approach include models implemented in the OASIS TIMES (Mekenyan et al. 2004, 2007; Serafimova et al. 2007), in CAESAR (Ferrari and Gini 2010) as well as some literature-based models not implemented in software (Purdy 1996).

The accuracy of QSARs for potency for both bacterial reverse mutation assay mutagenicity and rodent carcinogenicity (applicable only to toxic chemicals) is 30–70%, whereas the accuracy of classification models for discrimination between active and inactive chemicals is 70–100% depending on the model and dataset used. Usually accuracy of the models for carcinogenicity is relatively lower than what the bacterial reverse mutation assay gives. This is reasonable taking into account the complexity of the carcinogenicity endpoint, and the fact that models do not explicitly include ADME properties, which could be critical steps in the carcinogenic process. It has been argued that QSARs for carcinogenicity classification are of comparable performance to the gene mutation test in bacteria (Benigni et al. 2010).

Status of validation and/or standardisation: The validation process for a (Q)SAR model does not follow the validation procedures of in vitro test methods (<http://www.ecb.jrc.ec.europa.eu/qsar/background/>). It is a fast and an unofficial approach for characterising models and documenting them according to an internationally harmonised format, the QSAR Model Reporting Format (QMRF; <http://www.ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=QRF>). Since the usefulness of QSAR estimates is highly context dependent, there is no official acceptance or adoption process at EU or OECD levels.

The REACH legislation allows the use of QSAR models that are scientifically valid, applicable to the chemical of interest and that give results adequate for the regulatory purpose.

The validation procedure includes an assessment of model performance based on different statistical analyses (Eriksson et al. 2003). For models which are classifiers,

statistical parameters, such as accuracy (concordance), sensitivity and specificity, are used. For continuous (regression) models, a range of other parameters are typically used (e.g. the coefficient of determination, R^2 , and the standard error of the estimate, s). The ultimate proof of the predictivity of a QSAR is the demonstration that when applied to a new set of chemicals not used for the modelling (an independent test set), it predicts reliably their biological activity (Benigni and Bossa 2008).

The validation procedure also includes an assessment of the applicability domain of the model. The key is to understand whether it is appropriate to make a prediction for a given query chemical. Different chemometric approaches can be used to describe the applicability domain of a model and thus to assess model applicability. Some applicability domain methods are based on the structural similarity of the chemical of interest to the training set chemicals, whereas others are based on mechanistic similarity. The program AMBIT for example can evaluate this information (CEFIC 2010). Other approaches explore the possible use of other pieces of information for the applicability domain. For example, software based on the CAESAR model for carcinogenicity takes into consideration not only the chemical but also the toxicological information (CAESAR 2010). This method evaluates not only the input of the model (the chemical descriptors), but also the output, which is the toxicity property.

Unfortunately, there is no single and harmonised way of evaluating chemical similarity and defining applicability domains (Jaworska and Nikolova-Jeliazkova 2007), which means that the assessment of model applicability is not straightforward and needs to rely to some extent on expert judgement.

Field of application and limitations: (Q)SAR models are used by industry, in the upstream process, for a fast screening. More detailed information on the strengths and limitations of the different models is given elsewhere (Benfenati et al. 2009; Serafimova et al. 2010). Generally, the carcinogenicity models are not used for the final assessment.

It is interesting to notice that the results from different models may not agree, because they are based on different chemical information and rules. Indeed, any model is incomplete in its knowledge. Thus, the recommendation is to use more than one model and critically evaluate the results. Another possibility is to combine predictions of multiple models (Contrera et al. 2007). The results should be analysed considering the knowledge on the basis of the software. Indeed, several programs are quite transparent and show for instance the fragment which is supposed to trigger the toxic effect. When evaluating the reliability of a

prediction, it is important to critically evaluate not only the predicted outcome and the apparent predictive performance of the model, but also any supporting information that is available, such as whether the assumptions of the model have been fulfilled (e.g. the model is applicable to the substance being predicted) and information on the ability of the model to correctly predict suitable analogues of the chemical of interest.

Ongoing developments: There are some interesting perspectives for the integration of (Q)SAR with results of other tests. (Q)SAR models can offer advantages in the organisation and exploration of the data and information. This will be more powerful in case of the availability of a huge number of data, arising, for instance, from the ToxCast initiative (ToxCast 2010).

Read-across and grouping of chemicals Short description, scientific relevance and purpose: A chemical category is a group of chemicals whose physicochemical, human health, ecotoxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern, usually as a result of structural similarity (OECD 2007b). The grouping approach represents a move away from the traditional substance-by-substance evaluation to a more robust approach based on a family of related chemicals. Within a chemical category, data gaps may be filled by read-across, trend analysis and QSARs (van Leeuwen et al. 2009).

The OECD (2007b) guidance document on toxicological grouping of chemicals, which is based on the REACH guidance for grouping, proposes a stepwise approach for analogue read-across. The steps include (1) identifying potential analogues, (2) gathering data on these potential analogues, (3) evaluating the adequacy of data for each potential analogue, (4) constructing a matrix with available data for the target and analogue(s), (5) assessing the adequacy of the analogue(s) to fill the data gap and (6) documenting the entire process. The guidance also indicates the importance of comparing the physicochemical properties of the analogue and target chemicals as well as assessing the likely toxicokinetics of the substances, including the possibility that divergent metabolic pathways could be an important variable. Using the OECD guidance as a foundation, Wu et al. (2010) have recently published a framework for using similarity based on chemical structure, reactivity, and metabolic and physicochemical properties to specifically evaluate the suitability of analogues for use in read-across toxicological assessments.

Read-across interpolates or extrapolates the property of one or more compounds. For a given category endpoint, the category members are often related by a trend (e.g. increasing, decreasing or constant) in an effect, and a trend analysis can be carried out using a model based on the data

for the members of the category. Data gaps can also be filled by an external QSAR model, where the category under examination is a subcategory of the wider QSAR. All of these approaches can be performed in a qualitative or quantitative manner. In other words, using of a category approach means to extend the use of measured data to similar untested chemicals, and reliable estimates that are adequate for classification and labelling and/or risk assessment can be made without further testing.

Status of validation and/or standardisation: By its very nature, the grouping and read-across approach is an ad hoc, non-formalised approach based on a number of steps including expert choices. Thus, the term “validation” is not meaningful in this context. Instead, estimated properties obtained by the grouping and read-across approach need to be assessed in terms of their adequacy, and the justification needs to be clearly documented according to an accepted format (ECHA 2010). The critical issues of chemical category formation procedure are quality of the existing data for known chemicals and definitions of the similarity. The similarities may be based on the following: (1) structural features (e.g. common substructure, functional group, chemical elements), (2) physico-chemical, topological, geometrical, surface and quantum chemical properties, (3) behaviour (eco)toxicological response underpinned by a common mechanism of action and (4) toxicokinetics properties, including metabolic pathways.

At present, there are several software tools that can be used to build a category and fill data gaps related with genotoxicity and carcinogenicity (Serafimova et al. 2010). In version 2.0 of the OECD QSAR Toolbox, five mechanistically based profilers connected with genotoxicity and carcinogenicity are implemented (Enoch and Cronin 2010). In the software, 5 databases are included which contain genotoxic and carcinogenic experimental data. The Toolbox gives the possibility to form a category using also other criteria for similarity, including metabolism. Toxmatch (Patlewicz, et al. 2008) is another software tool that encodes several chemical similarity indices to facilitate the grouping of chemicals into categories and the application of read-across.

Compared to (Q)SAR methods, the experience on the use of these methods for carcinogenicity is limited. A limitation of these methods is that their reproducibility can be low, because the definition of the similar compounds and their number is not standardised. Different results are expected if the toxicity prediction is based on different compounds. More experience should be produced comparing results obtained from different users.

Field of application and limitations: Read-across is typically used when very similar compounds are present. Assessors rely on the property of these similar compounds. It is obvious that the reliability of this non-testing method

is highly related to toxicity values of the similar compounds. If the information is extracted from one or two chemicals, this information has to be very carefully checked. This applies to all non-testing methods, but in case of a large population of compounds, the presence of errors is less critical. Furthermore, interpolation should be preferred, compared to extrapolation.

Ongoing developments: Interesting perspectives exist in the development of more robust methods for similarity evaluation. For this, some of the tools above discussed for the applicability domain can be used for the evaluation of correctness of read-across and grouping.

Threshold of toxicologic concern (TTC) approach Cosmetics are typically mixtures of different ingredients added at varying levels, some of which are associated with very low exposure to consumers. For these ingredients, the TTC approach may offer a conservative, transparent and pragmatic way to assure safety. However, because of the conservative assumptions associated with TTC, it will be limited in its general applicability to cosmetics and will likely not be useful for ingredients used at higher levels or in products that involve higher exposures (e.g. body lotions).

Short description, scientific relevance and purpose. The TTC is a scientifically based approach to establish acceptable exposure limits when sufficient chemical-specific toxicologic information is lacking. It is a pragmatic risk assessment tool that relies on the broad grouping of chemicals based on structural features and then assumes that an untested chemical is potentially as toxic as the most toxic chemicals in the group. As a consequence, the TTC exposure limits are by design quite conservative. Furthermore, it is likely that if chemical-specific data were available, the risk assessment would support higher exposure levels. The intention of this approach is to provide a framework that minimises the time and effort spent on assessing low-level exposures by providing a means to develop scientifically supported exposure limits for these materials without the need to generate additional toxicity data. It is noted that TTC is different from some of the other alternative approaches described herein because the focus of TTC is on risk assessment (i.e. establishing an acceptable exposure limit) rather than being limited to hazard identification.

The origins of TTC as a risk assessment tool can be found in the US FDA's Threshold of Regulation (ToR), which was developed as a pragmatic way to assess the safety of low-level food packaging migrants (US FDA 1995). The ToR established an exposure level of 1.5 µg/day as being protective for chemicals lacking structural alerts for genotoxicity. This number was based on an analysis of the distribution of potencies of chemical carcinogens in the Carcinogen Potency Database (CPDB), which had 477 carcinogens in it at the time. Importantly, a re-analysis of a later update of the Gold's CPDB (1995)

that included more than 700 chemicals showed a similar distribution of cancer potencies (Cheeseman et al. 1999; Kroes et al. 2004). Although the ToR was based on an evaluation of cancer potencies, the exposure limit of 1.5 µg/day was not intended to be used with genotoxic carcinogens. This is because the Delaney Clause in US law prohibits the use of carcinogens as indirect food additives. Therefore, 1.5 µg/day was established as a limit that would not be used for chemicals with structural alerts or other reason for concern for genotoxicity, but would still be considered to be protective in the event that later testing revealed that the chemical did have some carcinogenic potential.

Status of validation and/or standardisation: The TTC methodology and scientific underpinnings continue to be expanded upon such that its utility and acceptance are growing. Since the initial work of the FDA on the ToR, the TTC methodology has been expanded into a tiered approach that has the potential for much broader applicability. Most notably, Kroes et al. (2004) describes the work of an Expert Group of ILSI Europe that culminated in the development of a decision tree that is now widely cited as providing the foundation for a tiered TTC approach. This publication describes a step-wise process in which it is first determined whether TTC is an appropriate tool (proteins, heavy metals and polyhalogenated-dibenzodioxins and related compounds have been so far excluded from use with TTC) and then follows a series of questions to determine the appropriate TTC tier. The initial step is the identification of high potency carcinogens that have currently been excluded from the TTC approach (aflatoxin-like, azoxy and N-nitroso compounds). After that, the chemical would be analysed for structural alerts for possible genotoxicity. Those with alerts would be assigned to the lowest TTC tier of 0.15 µg/day (an order of magnitude lower than the FDA's ToR). Organophosphates have then been assigned the next highest tier, followed by three higher tiers for non-genotoxic substances. These three tiers are based on the work of Munro et al. (1996b), who established a non-cancer database for TTC consisting of repeat-dose oral toxicity data of 613 substances. These substances were divided into three chemical classes of toxic potential on the basis of their structure using the decision tree of Cramer et al. (1978), and the distribution of NOELs was established for each of the three Cramer Classes. The 5th percentile NOEL was then calculated for each Cramer Class distribution, and an uncertainty factor of 100 was applied to establish human exposure thresholds of 1,800, 540 and 90 mg per person per day (30, 9 and 1.5 mg/kg bw/day) for the Cramer structural classes III, II and I. A new chemical lacking repeat-dose toxicity data could then be assigned a Cramer class based on structure, and the appropriate TTC value assigned.

Fields of application and limitations: Since its origin as a tool for food packaging materials in the mid-1990s, the acceptance and utility of TTC has been expanded such that it has been used extensively to assess food flavouring agents (JECFA 1996, 1997; EFSA 2004; Renwick 2004) and genotoxic impurities in pharmaceuticals (EMA 2006, 2008; Müller et al. 2006). The TTC decision tree has also been recommended as a tool to evaluate low-level exposures associated with personal and household products (Blackburn et al. 2005) and cosmetic ingredients and impurities in the absence of chemical-specific toxicology data (Kroes et al. 2007). Whereas the TTC databases are oral repeat-dose studies, cosmetic exposures are predominantly dermal. Therefore, in addition to considerations of the chemical domain, application of the TTC approach to cosmetics requires consideration of route-to-route extrapolation, including differences in absorption and first-pass metabolism. Kroes et al. (2007) published an analysis showing that the oral TTC values are in fact valid for use with dermal exposures. Furthermore, they recommended conservative default adjustment factors based on *in silico* prediction tools that could be used to estimate an absorbed dose following dermal exposure.

Ongoing developments: In addition to work ongoing to further expand the applicability and acceptance of TTC as a risk tool for cosmetic ingredients, other projects have aimed to expand the tool itself. For example, additional refinements have been recommended by Felter et al. (2009) that allow for the inclusion of genotoxicity data as a way to refine the TTC limit for chemicals that have structural alerts for genotoxicity and to support higher exposure limits for less-than-lifetime exposures. Also, work is ongoing to develop TTC as a tool to evaluate inhalation exposures (Carthew et al. 2009) from cosmetics and as a tool for safety assessment of sensitisation following dermal exposure (Safford 2008).

More recently, Bercu et al. (2010) proposed the use of TTC in combination with QSAR tools to establish safe levels for genotoxic impurities (GTIs) in drug substances. The single TTC limit of 0.15 µg/day is highly conservative and intended to be protective for the more potent end of the distribution of potencies for genotoxic chemicals (after excluding highly potent categories such as the N-nitroso carcinogens), and as such can be very restrictive in the development of new drug substances. To address this, Bercu et al. developed a tiered approach to use *in silico* tools to predict the cancer potency (TD50) of a compound based on its structure. Structure activity relationship (SAR) models were developed from the CPDB using two software packages: MultiCASE and VISDOM (Eli Lilly proprietary software). MultiCASE was used to predict a carcinogenic potency class, while VISDOM was used to predict a numerical TD50. For those compounds not categorised as

“potent” by MultiCASE, TD50 values were predicted by VISDOM that could then be used in establishing acceptable exposure levels. For those that were categorised as “potent”, the previously established TTC value of 0.15 µg/day would be used.

In vitro methods

Classical genotoxicity tests Short description, scientific relevance and purpose: Originally, *in vitro* genotoxicity tests are used to predict the intrinsic potential of substances to induce mutations. The rationale behind using genotoxicity tests for identifying potential carcinogens is that mutations and/or chromosomal aberrations are strongly associated with the carcinogenesis process. For this task, only *in vitro* genotoxicity tests which measure a mutation endpoint (gene or chromosomal mutation) are qualified: the gene mutation test in bacteria (OECD 471), the gene mutation test in mammalian cells (OECD 476), the chromosome aberration test (OECD 473) and the *in vitro* micronucleus test (OECD 487).

The tests rely on the fixation of initial DNA damage (DNA adducts or chromosomal damage) or damage to the cellular apparatus like the spindle figure into stable irreversible DNA modifications or changes in chromosome number. These modifications may result in the induction of diseases like cancer or genetic inheritable diseases. The tests are used to predict the potential of chemical substances to induce the former diseases.

Known users: Academics for mechanistic studies, all industries for screening purpose but also for regulatory application.

Status of validation and/or standardisation: With the exception of the *in vitro* micronucleus test (Corvi et al. 2008), none of the genotoxicity tests are formally validated but nonetheless established, scientifically accepted and used tests. For all the tests suggested, OECD guidelines exist.

Fields of application and limitations: The problem with *in vitro* genotoxicity tests, particularly for the tests measuring clastogenic effects, is the high number of misleading positives, i.e. positive test results for known non-carcinogens, as was discussed previously. Improvement of existing *in vitro* standard genotoxicity tests is under investigation. Preliminary data generated in a project sponsored by EC-VAM and predominantly the cosmetic industry show that misleading positive results can be reduced if: (1) *p53*-competent cells (e.g. human lymphocytes, TK6) are used instead of *p53*-compromised rodent cells (Fowler et al., *in press*), (2) cytotoxicity measures are based on proliferation during treatment instead of measures simply on cell count (Kirkland and Fowler 2010) and (3) the top concentration is reduced from 10 to 1 mM (Parry et al. 2010; Kirkland

and Fowler 2010). These modifications are in line with the OECD guidelines, except for the reduction of the top concentration which would need revision of the guidelines for in vitro genotoxicity testing.

Ongoing developments: The role of genotoxicity testing can be both qualitative (hazard assessment) and quantitative (risk assessment). A preliminary investigation on the applicability of in vivo genotoxicity tests to estimate cancer potency looked promising (see also the paragraph on in vivo genotoxicity test and Hernández et al., submitted). For a quantitative approach of in vitro genotoxicity tests, a foreseeable problem is the metrics comparison of the correlation, particularly how the dose of in vitro studies (in mM) translates to a dose in in vivo tests (mg/kg bw/day). For this reason, dose–response analysis of both in vitro and in vivo genotoxicity endpoints and carcinogenicity is essential. Unfortunately, dose–response analyses using sophisticated dose–response software such as PROAST (RIVM) or the BMDS (USEPA) have never been performed with in vitro genotoxicity tests. Given the promising results obtained between in vivo genotoxicity and carcinogenicity, it is worthwhile applying a similar approach to investigate whether in vitro genotoxicity tests are correlated to carcinogenic potency.

In vitro micronucleus test in 3D human reconstructed skin models (RSMN) Short description, scientific relevance and purpose: The micronucleus test in 3D human reconstructed skin models (RSMN) offers the potential for a more physiologically relevant approach especially regarding metabolic properties to test dermal exposure. It has been anticipated that these features of the reconstituted skin models could improve the predictive value of a genotoxicity assessment compared with that of existing in vitro tests and, therefore, could be used as a follow-up test in case of positive results from the standard in vitro genotoxicity testing battery (Maurici et al. 2005). Several 3D skin models are commercially available and are suitable for conducting such test, provided that sufficient cell proliferation is available.

Status of validation and/or standardisation: A RSMN protocol using the EpiDerm™ (MatTec Corporation, Ashland, MA, USA) model has been developed and evaluated with a variety of chemicals across three laboratories in the United States (Curren et al. 2006; Mun et al. 2009; Hu et al. 2009). A multilaboratory prevalidation study was initiated in 2007 and is sponsored and coordinated by the European Cosmetics Industry Association (COLIPA). This study aims at establishing the reliability of the method (Aardema et al. 2010) and at increasing the domain of chemicals tested for predictive capacity. Results generated so far show excellent inter- and intra-laboratory reproducibility and, therefore, suggest that the RSMN in

EpiDerm™ is a valuable in vitro method of dermally applied chemicals.

Fields of application and limitations: The test is aimed for use at chemicals for which there is dermal exposure. The RSMN test must be seen as an addition to the standard battery of in vitro genotoxicity tests. It will be important to demonstrate whether these tests have an equivalent sensitivity and a better specificity with regard to those of the standard in vitro micronucleus test.

Ongoing development: Research on the metabolic capacity of the test (Hu et al. 2010) and investigation of the utility of more complex models, such as full-thickness skin models, are ongoing.

In vitro comet assay in 3D human reconstructed skin models Short description, scientific relevance and purpose: The comet assay in 3D human reconstructed skin models is considered to be more relevant to evaluate the genotoxic potential of chemicals than when performed in cell cultures, because genotoxic effects can be evaluated under physiological conditions, especially regarding metabolic properties (Hu et al. 2010), and therefore be closer to the human situation than animal testing. This assay is a rapid and sensitive method to evaluate primary DNA damage, and it could be used as a follow-up test for chemicals that cause gene mutation in the in vitro standard tests (Maurici et al. 2005). Several 3D skin models are commercially available and are suitable for conducting such assay.

Status of validation and/or standardisation: Similarly to the RSMN test, a protocol using the EpiDerm™ model has been developed for the comet assay in 3D human reconstructed skin models and is being optimised and evaluated across three laboratories in the EU and USA. This study which aims at establishing the reliability of the method and at increasing the domain of chemicals tested for predictive capacity was initiated in 2007 and is coordinated by COLIPA and sponsored by COLIPA and ECVAM.

Fields of application and limitations: The test is aimed for use at chemicals for which there is dermal exposure. It must be seen as an addition to the standard battery of in vitro genotoxicity tests. Being the endpoint sensitive to DNA damage, it is crucial that the quality of the tissues and good shipping conditions are ensured.

Ongoing development: Research on the metabolic capacity of the assay (Hu et al. 2010) and investigation of the utility of more complex models, such as full-thickness skin models, are ongoing. Moreover, application of the comet assay to freshly obtained human skin tissue that is generally obtained following cosmetic surgery is under investigation.

GreenScreen HC assay Short description, scientific relevance and purpose: The GreenScreen HC (Gentronix Ltd,

Manchester, UK) is a commercially available assay for genotoxicity testing, using human lymphoblastoid TK6 cells transfected with the *GADD45a* (*growth arrest and DNA damage*) gene linked to a green fluorescent protein (GFP) reporter (Hastwell et al. 2006). This assay is based on the upregulation of *GADD45a*-GFP transcription and the subsequent increase in fluorescence, in response to genome damage and genotoxic stress. The test can be performed with or without metabolic activation by S9-liver fraction.

Status of validation and/or standardisation: Standard protocols have been developed for both methods, without (Hastwell et al. 2006) or with (Jagger et al. 2009) metabolic activation, and their transferability, within-laboratory reproducibility (Hastwell et al. 2006; Jagger et al. 2009) and between-laboratory reproducibility (Billinton et al. 2008, 2010) have been evaluated.

Fields of application and limitations: This test is used by the pharmaceutical industry as early screening tool in drug discovery. However, most pharmaceutical companies are still investigating the utility of the screen in their strategies and how to interpret the data for internal decision-making.

Some technical aspects have also to be taken into account for the conduct of the test: the protocol in the absence of metabolic activation only requires the use of a microplate spectrophotometer and is compatible with high-throughput screening, whereas the accessibility and the automation of the S9-protocol are both limited by the necessity of a flow cytometer to avoid interference with the light-absorbing and fluorescent properties of S9-particulates.

Ongoing development: A variant of the S9-protocol has been developed, which was adapted to microplate readers by the use of a fluorescent cell stain and fluorescence (instead of absorbance) measurement to estimate cell number. Although flow cytometry remains the most sensitive method, this variant is more suitable for non-flow cytometer users and for high-throughput screening.

The BlueScreen HC is a new assay under development that uses the same *GADD45a* reporter gene as the Green-Screen HC assay but linked to *Gaussia luciferase* gene, which leads to a greater signal-to-noise ratio than with GFP. Moreover, it has full compatibility with S9-liver fraction use and thus with high-throughput screening capability.

Hens egg test for micronucleus induction (HET-MN)

Short description, scientific relevance and purpose: Another promising system as a follow-up for in vitro positive for cosmetic ingredient is the hens' egg test for micronucleus induction (HET-MN; Wolf et al. 2008). The HET-MN combines the use of the commonly accepted genetic endpoint "formation of micronuclei" with the well-

characterised and complex model of the incubated hen's egg, which enables metabolic activation, elimination and excretion of xenobiotics, including those that are mutagens or promutagens. The assay procedure is in line with demands for animal protection.

Status of validation and/or standardisation: A prevalidation study is planned starting in September 2010 with at least three participating laboratories investigating the transferability and intra-laboratory reproducibility. Results of this study will most probably be available in 2012.

Fields of application and limitations: At present, the HET-MN is not frequently used. Only few laboratories have established this test for screening purposes. Studies on metabolism indicate that certain important phase I and II enzymes are active and, therefore, the detection of liver mutagens is possible. Up to now, the transferability and intra-laboratory reproducibility is not provided.

Ongoing developments: An improvement may be the inclusion of flow cytometric analysis where higher cell numbers can be evaluated in a shorter time and which could improve the sensitivity of the assay as the sample size can be dramatically increased.

Cell transformation assay Short description, scientific

relevance and purpose: Mammalian cell culture systems may be used to detect phenotypic changes in vitro induced by chemical substances associated with malignant transformation in vivo. Widely used cells include SHE, C3H10T_{1/2}, Balb/3T3 and Bhas 42 cells. The tests rely on changes in cell colony morphology and monolayer focus formation. Less widely used systems exist which detect other physiological or morphological changes in cells following exposure to carcinogenic chemicals. Cytotoxicity is determined by measuring the effect of the test material on colony-forming abilities (cloning efficiency) or growth rates of the cultures.

Status of validation and/or standardisation: In 2007 the OECD published a Detailed Review Paper (DRP31) aiming at reviewing all the available data on the 3 main protocols for cell transformation assays and concluded that the performance of the assays using SHE and Balb/c 3T3 cells was sufficiently adequate (OECD 2007e) and should be developed into OECD test guidelines. A prevalidation study with SHE (pH 6.7 and 7.0) and Balb/c 3T3 cells was organised by ECVAM to address issues of standardisation of the protocols, transferability and reproducibility. The experimental work was finished in 2009. The data demonstrated that the SHE protocols and the assays system themselves are transferable between laboratories and are reproducible within- and between laboratories. For the Balb/c 3T3 method, an improved protocol has been developed, which allowed to obtain reproducible results. Further testing of this improved protocol is recommended

in order to confirm its robustness (Vanparys et al. 2010). Overall, these results in combination with the extensive database summarised in the OECD DRP31 (OECD 2007e) will support the development of the OECD test guidelines for the assessment of carcinogenicity potential. This ongoing work should progress in the coming 3 years.

Fields of application and limitations: The in vitro cell transformation assays have been established in order to predict tumourigenicity (DiPaolo et al. 1969; Isfort et al. 1996; Matthews et al. 1993). Some of the test systems are capable of detecting tumour promoters (Rivedal and Sanner 1982). Some cell types and substances may require an appropriate external metabolic activation system. When primary cells are used that possess intrinsic metabolic activity, additional metabolic activation is not needed. The scoring of transformed colonies and foci may require some training and experience.

The cell transformation assays are currently used for clarification of in vitro positive results from genotoxicity assays to be used in the weight of evidence assessment. Data generated by cell transformation assays can be useful where genotoxicity data for a certain substance class have limited predictive capacity (e.g. aromatic amines), for investigation of compounds with structural alerts for carcinogenicity or to demonstrate differences or similarities across a chemical category. Also the tumour-promoting activity of chemicals can be investigated by the cell transformation assays.

Known users: Academics, pharmaceutical and agro-chemical industry for screening purpose, cosmetic industry and chemical industry also for regulatory application. Academia is using it for mechanistic studies.

Ongoing developments: Certain improvements for investigating the transformed phenotype have been proposed. Transformed colonies can be detected by discrimination of the transformation phenotype by using ATR-FTIR spectroscopy (Walsh et al. 2009), by image analysis (Urani et al. 2009) or by the inclusion of molecular biomarkers (Poth et al. 2007). The technical performance of the SHE assay has been improved by avoiding the use of X-ray irradiated feeder layers (Pant et al. 2008). Systems biology is included for mechanistic investigation of cellular transformation (Ao et al. 2010; Rohrbeck et al. 2010) and also the throughput has been increased by using soft agar colony screening (Thierbach and Steinberg 2009) and Bhas 42 96-well plate method (Ohmori et al. 2005).

In vitro toxicogenomics Short description, scientific relevance and purpose: Since the introduction of genomic technologies circa 10 years ago, their application in toxicology, toxicogenomics, has developed enormously (Ellinger-Ziegelbauer et al. 2009a; Guyton et al. 2009; Waters et al. 2010). The unbiased analyses of global

perturbations by chemicals in cells and organisms at the level of genes, transcripts, proteins and metabolites, in combination with powerful bioinformatic tools, provides an unprecedented wealth of information about the molecular processes and mechanisms that can be affected. This knowledge can be used for elucidating the mode of action of compounds, prediction of toxic properties, cross-species and in vitro-in vivo comparison, and even in epidemiological settings for assessment of exposure and (adverse) effects in humans. Predominantly, transcriptomics (gene expression analyses at the level of mRNA) has received most attention and has proven to be promising. For in vitro hazard assessment in the area of genotoxicity and carcinogenicity TK6, HepG2 and primary liver cells are mostly used. These toxicogenomics approaches reach 80–90% accuracy (Ellinger-Ziegelbauer et al. 2009a; Li et al. 2007; Tsujimura et al. 2006; Le Fevre et al. 2007; Hu et al. 2004; Mathijs et al. 2010) for predicting in vivo toxicity in rodents although the number of chemicals is still limited and may not represent the full spectrum of toxins.

Known users: Academia is using it for mechanistic studies. Pharmaceutical and cosmetic industry in screening purposes only.

Status of validation and/or standardisation: No formal validation under the guidance of ECVAM or a similar organisation has been performed although the technology has been extensively evaluated by the MicroArray Quality Consortium (MAQC; Guo et al. 2006; MAQC Consortium 2006). For some tests based on gene expression analyses, standard protocols are being developed and optimised (see “Ongoing developments”). This ongoing work will progress in the coming 3 years; depending on the results and conclusions of these studies, some tests might be ready to enter prevalidation.

Fields of application and limitations: Various tests based on gene expression analyses can be foreseen in the near future for screening purposes and labelling of compounds and, thus for hazard assessment only. Tests for genotoxicity and carcinogenicity in general, or for specific mechanisms therein, are under development (Mathijs et al. 2010). The transcriptomic biomarkers will be complex, consisting of profiles for multiple genes. As many genes have been annotated with respect to their function and sometimes to toxicological pathways (e.g. the DNA damage pathway), this will provide mechanistic information as well. Since cell types and substances may require an appropriate external metabolic activation system, in general, cells derived from liver that possess intrinsic metabolic activity are used making additional metabolic activation redundant. The limitations are many, such as risk assessment is problematic, each assay focuses on a specific aspect of genotoxicity or carcinogenicity, limited public accessibility of raw data, the function of many genes in the

prediction sets are not understood, lack of uniformity in study design (e.g. cell lines, dose setting criteria, time points, repeats, etc.) and bioinformatic analyses, and the requirement of expensive equipment and specialised staff.

Ongoing developments: Recently, a multilaboratory project coordinated by the Health and Environmental Sciences Institute (ILSI-HESI) demonstrated that expression analysis by RT-PCR of a relevant gene set derived from -omics data, is capable of distinguishing compounds that are DNA reactive genotoxins from those that are non-DNA reactive genotoxins (Ellinger-Ziegelbauer et al. 2009b). RT-PCR provides a cheaper and faster test for gene expression profiling, when limited to relatively small gene sets. Furthermore, as part of the EU-funded project “carcinoGENOMICS” several *in vitro* models for liver, lung and kidney are tested (cell lines, stem cell-derived hepatocytes) and compared and certain aspects related to reliability of the tests will be addressed in 2010–2011 (<http://www.carcinogenomics.eu>).

In vivo methods (reduction/refinement)

Since no complete replacement methods are to date available in the area of carcinogenicity, we considered also reduction and refinement methods. The use of the described *in vivo* assays can, in the absence of a validated alternative to carcinogenicity testing, serve as a reduction approach as they use at least 90% less animals. As alternatives, we mention only those tests which are considered as predictive tests for carcinogenicity and those which continue to undergo further development. Rather old approaches like the liver foci assay and the neonatal mouse assay are for this reason not described further.

In vivo genotoxicity tests For the same reason as for *in vitro* tests, also *in vivo* genotoxicity tests are a tool to predict cancer risk. Contrary to the *in vitro* tests, the *in vivo* tests do address ADME. As for the *in vitro* genotoxicity tests, also for *in vivo* genotoxicity tests, the specificity and predictivity have to be at a level that the prediction of carcinogenicity is justified. This again may lead to modifications of test protocols, i.e. species, (top) doses, among others. But then, a positive result in an *in vivo* genotoxicity test may point to a carcinogenic potency of the compound under investigation and further carcinogenicity testing may not be needed. As for the *in vitro* tests, predominantly tests which measure irreversible genotoxic damage may be considered: the chromosome aberration test (OECD 475), the micronucleus test (OECD 474) and the gene mutation test with transgenic animals (OECD guideline in prep). However, as an exception, the comet assay (validation ongoing, OECD guideline foreseen) is considered a useful test, since it covers both the gene mutation and the

chromosome aberration endpoint. Furthermore, if the most modern approaches are used and the flexibility the OECD guidelines provided are fully utilised, these tests can be performed using even less animals compared to a standard guideline test (50% or more reduction in animal numbers; Pfuhler et al. 2009).

The role of genotoxicity testing can be both qualitative (hazard assessment) and quantitative (risk assessment). The finding of a linear relationship between the lowest effective dose (LED) for *in vivo* genotoxicity and the carcinogen dose descriptor T25 is of importance (Sanner and Dybing 2005). It was found for 34 carcinogens studied which covered a potency range of 10,000, that the median of the ratio LED/T25 was equal to 1.05 and that for 90% of the substances the numerical value of LED was similar to the numerical value of T25 within a factor of less than 5–10. The results suggest that if further evaluated, LED for *in vivo* genotoxicity may be used in a semi-quantitative method for risk assessment of mutagens without a long-term study.

The above results are further supported by a preliminary investigation on the applicability of genotoxicity tests to estimate cancer potency undertaken in the RIVM using the benchmark dose approach. Positive correlations between *in vivo* genotoxicity (micronucleus test and transgenic rodent mutation test) and carcinogenic potency were found (Hernández et al., submitted). Dose–response analyses using sophisticated dose–response software such as PROAST (RIVM) or the BMDS (USEPA) were used. The results suggest that *in vivo* genotoxicity tests may be used to estimate carcinogenic potency.

Transgenic mouse models Short-term tests with transgenic mouse models ($p53^{+/-}$, $rasH2$, Tg.AC, $Xpa^{-/-}$ and $Xpa^{-/-}p53^{+/-}$) are a good alternative to the classical two-year cancer bioassay (Ashby 2001). The rationale for using transgenic mice in regulatory carcinogenicity testing is that transgenic mouse models may be more sensitive predictors of carcinogenic risk to humans. Indeed these transgenic mouse models had a reduced tumour latency period (6–9 months) to chemically induced tumours (Marx 2003). The increased sensitivity to tumour formation in transgenic mouse models is primarily due to modifications in the mouse genome by either removing or adding specific genetic material (Tennant et al. 1995, 1999). Although not a complete replacement to the rodent 2-year cancer bioassay, transgenic mouse models are a refinement and result in a significant reduction in the use of experimental animals.

Several studies (ILSI/HESI ACT 2001; Eastin et al. 1998; Bucher 1998; Pritchard et al. 2003; de Vries et al. 2004) demonstrate that in all transgenic models, a limited number of animals 20–25 animals/sex/treatment group can be used and that an exposure of 6–9 months is sufficient.

However, wild-type animals should be included in test battery to demonstrate that no genetic drift may affect interpretation of the results. Transgenic mouse models showed a high specificity given that all non-carcinogens tested gave negative results in all 5 transgenic models. These findings provide evidence against “oversensitivity” concerns associated with transgenic mouse models due to modifications in cancer-related genes. Transgenic models were able to discriminate not only between carcinogens and non-carcinogens but even between rodent carcinogens and putative human non-carcinogens to a high degree of accuracy.

In vivo toxicogenomics In the paragraph on *in vitro* toxicogenomics, an introduction is given on relevance of genomics technologies and their application in toxicology. Also in case of *in vivo* toxicogenomics, gene expression analysis (transcriptomics) has been developed further. Most *in vivo* toxicogenomics studies on assessment of carcinogenicity, focus, but do not limit, on short-term rat studies and non-genotoxic hepatocarcinogenicity. These toxicogenomics approaches reach 80–90% accuracy (Ellinger-Ziegelbauer et al. 2008; Nie et al. 2006; Fielden et al. 2007; Stemmer et al. 2007; Nioi et al. 2008; Uehara et al. 2008; Jonker et al. 2009) for predicting rodent carcinogenicity which is in the same range as for *in vitro* toxicogenomics. Pharmaceutical and sometimes the chemical industries are the main users for screening purposes. In rare cases, the pharmaceutical industry also uses *in vivo* toxicogenomics for mechanistic purposes.

Various assays based on gene expression analyses can be foreseen for the near future, which can be used for screening/prioritisation purposes and for labelling of compounds. These assays can also be helpful for the understanding of modes of action. Especially assays for non-genotoxic hepato-carcinogenicity are under development (Ellinger-Ziegelbauer et al. 2008; Nie et al. 2006; Fielden et al. 2007; Stemmer et al. 2007; Nioi et al. 2008; Uehara et al. 2008; Jonker et al. 2009). Like for *in vitro* toxicogenomics, the limitations are many, such as quantitative risk assessment is in its infancy, strong focus on non-genotoxic carcinogens (mainly pharmaceuticals) and on the liver as target organ, limited public accessibility of raw data, the function of many genes in the prediction sets are not understood, lack of uniformity in study design (e.g. rodent species and strain, dose setting criteria, time points, repeats, etc.) and bioinformatic analyses, and the requirement of expensive equipment and specialised staff.

The *in vivo* toxicogenomics assays could be very helpful for hazard assessment and by that may lead to a reduction in the number of bioassays and the number of animals in the remaining *in vivo* tests. The number of animals required for toxicogenomics based assays is at least 10-fold

smaller than for the rodent bioassay, and the exposure periods last maximally 4 weeks instead of 2 years.

So far, no formal validation of a method has been performed. A gene expression profile for rat hepatocarcinogenicity is being investigated for some aspects of reliability. This ongoing work will progress in the coming 3 years; depending on the results and conclusions of these studies, some tests may be ready to enter pre-validation. The Predictive Safety Testing Consortium of the Critical Path Institute evaluated the predictivity of two published hepatic gene expression signatures when sharing the data (Fielden et al. 2008). Based on the results, a QPCR-based signature has been derived and is currently evaluated for inter-laboratory precision, sensitivity and specificity, time-dependency and non-genotoxic versus genotoxic mechanisms.

Identified areas with no alternative methods available and related scientific/technical difficulties

This report has highlighted a number of *in vivo* studies historically used in the safety evaluation of cosmetics with respect to carcinogenicity. For carcinogenicity, the gold standard of a 2-year bioassay is not commonly used, but several shorter-term studies are conducted, including *in vitro* and *in vivo* genotoxicity studies, repeat-dose studies and other mechanistic studies used in the safety assessment of non-genotoxic compounds. Therefore, the animal testing ban has implications well beyond the 2-year bioassay for the evaluation of cancer hazard and risk. The first challenge is for *in vitro* genotoxic compounds, where the strategy until 2009 allowed for *in vivo* genotoxicity testing which was the standard tool to clarify the relevance of *in vitro* positives. The difficulties with the *in vitro* tests have been described by Kirkland et al. (2007), and work is ongoing to improve these assays. The second challenge is for genotoxic carcinogens. In case nothing would change in available test methods/strategies, we can only rely on *in vitro* tests which test for intrinsic properties with a substantial rate of misleading positives in the classical tests. Consequently, the carcinogenic potential of substances will not be characterised with the same level of certainty as today.

The ban on repeat-dose toxicity testing raises new questions whether safety can be assured specifically for non-genotoxic chemicals. It is noted that this is the same challenge posed to those charged with developing alternatives for target organ toxicity. In the past, the safety assessment for non-genotoxic chemicals has been based on identification of a NO(A)EL from repeat-dose toxicity studies. If in the future as it will no longer be possible to conduct those studies for cosmetics ingredients, there will be a lack of information on the non-genotoxic endpoint. Given that some non-genotoxic carcinogens are known

human carcinogens and the potential hazard associated with them, there is a need for the development of alternative methods for the detection and risk assessment of non-genotoxic carcinogens. Ideally, these alternatives should include individual endpoints that are typically targeted by non-genotoxic carcinogens, such as the induction of oxidative stress and the inhibition of gap junction intercellular communication. Today, however, these tests are research methods, primarily used for evaluating mechanism and cannot currently be used to predict whether a chemical will be a carcinogen or what its potency would be. Cell transformation assays have been developed as a tool to identify both genotoxic and non-genotoxic carcinogens. The development of the OECD guideline for these assays is in progress. Toxicogenomics is an emerging area and also offers promises for the detection of non-genotoxic carcinogens, but still stands in its infancy. The challenges of extrapolation from *in vitro* to *in vivo* are also that most *in vitro* studies are limited to use in hazard identification and cannot yet be used in risk assessment. These have been described by Blaauboer (2010) and are also covered by the working group which develops the alternatives for toxicokinetics.

In silico methods such as QSAR have proven successful at predicting genotoxic potential and rationalising the chemical basis in terms of DNA reactivity. Such QSARs can be as reliable and informative as the gene mutation test in bacteria (Ames test), provided that their predictive algorithms and applicability domains are well characterised (Benigni et al. 2010). More research is needed to understand how the applicability domains relate to the chemical classes used in cosmetics. An advantage of the QSAR approach is that the models can be tuned to meet user-defined performance criteria such as low false positive or negative rates, depending on their foreseen use in a testing strategy. Relatively, few QSARs are available for non-genotoxic carcinogenicity and carcinogenic potency, and this represents a knowledge/development gap. The category and read-across approach provides a promising means of filling qualitative and quantitative data gaps (van Leeuwen et al. 2009). However, specialised knowledge and tools are needed to build the category and draw conclusions on the adequacy of the read-across. It is not possible to validate this approach *a priori*. In addition to being a stand-alone approach, read-across can also be used to add confidence to a prediction generated by QSAR. Additional confidence can be provided by *in vitro* data. At present, it is suggested to apply QSAR and read-across within the context of a Weight of Evidence or TTC approach. This implies the use of multiple QSARs in combination with each other (e.g. Matthews et al. 2008) and if possible with *in vitro* tests (e.g. Peer Consultation on Health Canada

Draft Weight of Evidence Framework for Genotoxic Carcinogenicity 2005).

In order to test a cosmetic ingredient for the evaluation of its carcinogenic potential, COLIPA (Pfuhler et al. 2010) has proposed a tiered testing strategy focused on genotoxic carcinogens for cosmetic ingredients for beyond 2013. However, this strategy does not take into account non-genotoxic carcinogens. No strategy is currently in place to detect non-genotoxic carcinogens.

Conclusions

The process of carcinogenesis is recognised as a multistep/multistage process from the transition of normal cells into cancer cells via a sequence of stages and complex biological interactions, strongly influenced by factors such as genetics, age, diet, environment, hormonal balance, etc. It is also recognised that there are many different modes of action that can contribute to the carcinogenic process and that even for one chemical the mode of action can be different in different target organs, or in different species. Despite best efforts, at present the modelling of such complex adverse effects cannot fully be accomplished by the use of non-animal tests.

As *in vivo* testing is no longer possible, the safety of many potential new cosmetic ingredients will not be substantiated and will therefore not be allowed to be marketed.

For genotoxic chemicals, a number of *in vitro* genotoxicity tests that are currently used to screen chemicals for activity that is considered to be predictive of potential carcinogenicity are available. While these tests have good sensitivity, some (especially the *in vitro* mammalian cell tests) have a high rate of misleading positives. Prior to 2009, a positive finding in an *in vitro* study was commonly followed by an *in vivo* study to clarify the *in vitro* results. Indeed, the vast majority of compounds tested *in vivo* was negative. Because of the 2009 ban on *in vivo* genotoxicity testing, the situation is now problematic and many potential cosmetic ingredients may be lost because of an inability to clarify misleading positive results from an *in vitro* genotoxicity test. Work is ongoing to improve these *in vitro* tests.

Cell transformation assays are to date the only *in vitro* tests that have reached a certain level of standardisation and have the potential to detect both genotoxic and non-genotoxic carcinogens. However, at the moment, these assays cannot be considered as a stand-alone solution to detect human carcinogens, but have the potential to contribute to a weight of evidence approach. Importantly, these assays are currently useful only in the hazard identification of carcinogens; there are no methods yet to use data from these tests to support a risk assessment.

For non-genotoxic chemicals, the standard approach for risk assessment (globally and for all sectors, not limited to cosmetics) has been to assume a threshold for cosmetic ingredients in general based on the results of repeated dose toxicity studies. A NOAEL from a repeat-dose toxicity study, along with appropriate conservative safety factors, has been used to estimate a risk assessment for these chemicals, including the risk for carcinogenicity. When repeat-dose toxicity testing is banned in 2013, methods for quantitative assessment of non-genotoxic carcinogenic risks will be limited to tools such as read-across, (Q)SAR and TTC (pending acceptance by the SCCS). Currently, due to limited experience, QSAR and read-across methods are better suited for use within a weight of evidence approach rather than as stand-alone methods. Because of limitations, the TTC does default to very conservative assumptions that further limit the utility of these approaches confirming safety during the development of new cosmetic ingredients. Indeed, consumer exposure to many ingredients is too high for TTC to be useful in many cases. Rather, it is expected to be useful only for contaminants and/or low-level ingredients associated with very low consumer exposures.

Although many *in vitro* short-term tests are available beyond the standard *in vitro* genotoxicity tests to support conclusions on cancer hazard identification, the *in vitro* short-term tests will not be sufficient to fully replace the animal tests needed to perform risk assessment for carcinogenicity for cosmetic ingredients. However, for some chemical classes, the available non-animal methods might be sufficient to rule out carcinogenic potential in weight of evidence approach.

Taking into consideration the present state of the art of the non-animal methods, the experts were unable to suggest a timeline for full replacement of animal tests currently needed to fully evaluate carcinogenic risks of chemicals. Full replacement is expected to extend past 2013.

Reproductive toxicity

Executive summary

In the last decades, significant efforts have been undertaken to develop alternative methods to assess reproductive toxicity. However, despite the impressive number of alternative tests that have been published and are listed in this report, the majority of these tests have not yet gained regulatory acceptance. There are a number of reasons for the relatively slow progress in the implementation of alternative methods for reproductive toxicity safety evaluations, these include the lengthy research and development phase, a lack of understanding of the mode of actions

of reproductive toxicants and the huge number of physiological mechanisms involved in mammalian reproduction which can be affected by xenobiotics. Among the various stages in the reproductive cycle, embryo-foetal development is considered as one of the most critical steps. Substantial effort has been spent in the development of promising *in vitro* assays, such as the zebrafish embryo test and pluripotent embryonic stem cell models, to allow for the detection of the teratogenic potential of substances. However, besides their current role as mechanistic support and screening tools, the role of alternative methods as part of integrated testing strategies for regulatory toxicity evaluations has to be defined further.

The complexity of mammalian reproduction requires integrated testing strategies to fulfil all needs for hazard identification and risk assessment. A promising way forward is the use of recently established comprehensive databases in which toxicological information derived from standardised animal experimentations is collected. These databases will allow for the identification of the most sensitive targets of reproductive toxicants. This priority setting of sensitive endpoints is the first step to obtain a detailed understanding of the toxicological relevance of the *in vitro* tests described in this report and how they can be used in integrated testing strategies. Furthermore, this mapping exercise will also support the identification of information gaps where further efforts in test development are necessary to design specific alternative methods covering identified sensitive endpoints.

According to the actual Cosmetics Directive 76/768/EEC and its 7th amendment, only validated alternatives leading to full replacement of animal experiments are of relevance for safety evaluations of cosmetic and their ingredients. Regardless, the retrospective analysis of available *in vivo* data allowing the detection of the most sensitive endpoints, the definition of a toolbox of alternative methods as well as the eventual need to develop additional alternatives to cover the missing building blocks in the testing strategy will need more than 10 years to complete.

Introduction

Complexity of the reproductive cycle

Reproductive toxicity refers to a wide variety of toxicological effects that may occur in different phases within the reproductive cycle (Fig. 10). This includes effects on fertility, sexual behaviour, embryo implantation, embryonic/foetal development, parturition, postnatal adaptation, and subsequent growth and development into sexual maturity. An enormous variety of mechanisms at the molecular, cellular and tissue levels cooperate in a concerted and

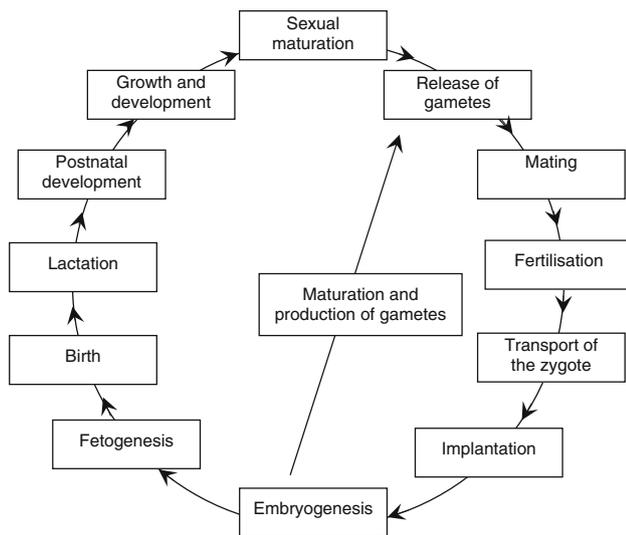


Fig. 10 The main stages in the mammalian reproductive cycle

genetically programmed way to regulate these processes. The sensitivity to chemical insults may differ extensively between processes. In addition, different temporal windows of sensitivity have been observed for different processes. As an example, neural tube closure occurs early in pregnancy, and most effects on this process can only be determined after exposure during this critical period of time.

Alternatives for reproductive toxicity testing

Over the last two decades, a wealth of *ex vivo* and *in vitro* assays have been proposed as alternative test systems for testing toxic effects on the various processes in reproduction and development. Individual *in vitro* models are reductionistic in nature and are therefore unable to cover all aspects of the reproductive cycle since reproduction requires a complex interplay of integrated functions (Piersma 2006). However, parts of the reproductive cycle can be mimicked by *in vitro* systems, and it is conceivable that a panel of well-designed and validated *in vitro* tests could replace a substantial proportion of *in vivo* testing procedures. This chapter gives an inventory of the current state of development of alternative test systems for reproductive toxicity hazard assessment.

Although not applicable for cosmetic ingredients, refinement and reduction of animal studies is a more feasible goal than replacement, one example being the current OECD activity towards an extended-1-generation study protocol, which, if it would replace the current 2-generation study, would reduce animal use by roughly 40% in each study (Cooper et al. 2006). The addition of relevant parameters to this novel study protocol represents a good example of refined testing.

Information requirements for the safety assessment of cosmetic

Within the EU the safety of cosmetic products is regulated by the Cosmetics Product Directive 76/768/EEC (EU 1976) which will be replaced stepwise by the new EU Cosmetics Regulation 1223/2009. According to Article 2 of Directive 76/768/EEC, a “cosmetic product put on the market must not cause damage to human health when applied under normal or reasonably foreseeable conditions of use”. In addition, Article 7a of the same Directive states that the safety evaluation of a finished product should be based on the general toxicological profile, the chemical structure and the level of exposure of each ingredient. This implies that a quantitative risk assessment is required for each single ingredient of a cosmetic product.

Being responsible for the safety of its cosmetic product, the producer assigns a qualified safety assessor who performs a risk assessment based on the data of all ingredients used. It must be emphasised, however, that not all cosmetic ingredients have been subject to a pre-market approval involving extended toxicological data requirements. In fact, the latter is mainly reserved for those ingredients listed in the positive lists of the Cosmetics Directive, such as colorants (Annex IV), preservatives (Annex VI), UV filters (Annex VII) and other substances which might involve a potential health risk for the consumer, including hair dyes (Annex III) (Pauwels and Rogiers 2010). These ingredients are evaluated by the Scientific Committee on Consumer Safety (SCCS, former SCCP), and details of their data requirements can be retrieved from the SCCP Notes of Guidance (Scientific Committee on Consumer Products (SCCP 2006)).

The requested comprehensive dossier to be submitted to the SCCS includes data on acute toxicity (if available), dermal and mucous membrane irritation, dermal penetration, skin sensitisation, repeated dose toxicity, genotoxicity and phototoxicity (if the cosmetic product is intended to be used on sunlight-exposed skin). Further, it is stated that when considerable oral intake is expected, or when dermal penetration data suggest a significant systemic absorption, information on toxicokinetics, carcinogenicity and reproductive toxicity “may become necessary”. Additional recommendations on specific *in vivo* or *in vitro* reproductive toxicity studies to be submitted with a dossier are not described in the Notes on Guidance. From the SCCS/SCCP opinions published within recent years (2000–2009) (http://www.ec.europa.eu/health/ph_risk/committees/04_sccs/sccs_opinions_en.htm), it can be concluded that in most cases an *in vivo* developmental toxicity study in the rat (OECD TG 414)—submitted by the manufacturer as the only study on reproductive toxicity—was considered sufficient by the SCCS as the minimum requirement. In only a few cases,

additional data from a 1- or 2-generation study (OECD TG 415 and 416) were included in a dossier (Pauwels and Rogiers, in preparation).

For substances, which are not listed in one of the Annexes of the Cosmetic Directive, data on reproductive toxicity are not explicitly asked for in the Notes of Guidance. However, according to Regulation EC 1223/2009 (recast of the Cosmetic Directive), the toxicological profile of each cosmetic ingredient must be assessed by a responsible person taking into account all significant routes of absorption. This safety evaluation includes systemic effects and the calculation of a Margin of Safety (MoS) for each ingredient. As a cosmetic product put on the market must not cause damage to human health when applied under normal or reasonable foreseeable conditions for use, possible effects on reproduction and development must be considered for each cosmetic ingredient. Some indications of adverse effects on the fertility could be obtained, e.g. from repeated dose toxicity studies, if available (e.g. histopathological effects on reproductive organs, effects on the endocrine system).

Inventory of animal test methods currently used for the evaluation of developmental and reproductive toxicity

The Organisation for Economic Co-operation and Development (OECD) have been producing highly standardised and internationally harmonised test guidelines to be used for the regulatory toxicological evaluation of different products, including industrial chemicals, agrochemicals and cosmetics. In addition, it has recently prepared a Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment (OECD 2008b).

The older guidelines for the evaluation of developmental and reproductive toxicity have been designed to include all known endpoints according to expert judgement. Recent guidelines (e.g. Uterotrophic and Hershberger) have undergone a validation exercise. Extrapolation of animal data to humans is complex due to inter-species differences (Hurtt et al. 2003). These species and strain differences include variations in the absorption, distribution, metabolism and excretion of chemicals; in placental structure, permeability and blood flows (Schroder 1995); and in the genetic backgrounds of different species (Kawakami et al. 2006).

The US FDA published a report detailing the responses of the mice, rats, rabbits, hamsters and monkeys to 38 known human teratogens in which the mean percentage of correct positives from any one of these species was 60% (Frankos 1985). Hurtt et al. (2003) found that positive predictivity of one animal species to teratogenic effects in either rat, mouse or rabbit was around 60% for 105

veterinary pharmaceuticals. Bailey (2005) examined the data for 11 groups of known human teratogens across 12 animal species and found huge variability in positive predictability, with a mean of 61%.

As previously expressed, for developmental and reproductive toxicity evaluation of cosmetics, only TG 414 is required, complemented in a few cases with 415 and 416. However, for informative reasons, the main guidelines are summarised.

OECD test guideline 414: prenatal development toxicity study for the testing of chemicals (OECD 2001a)

- Studies the effects of prenatal exposure on the pregnant animal and on the developing organism; this may include assessment of maternal effects as well as death, structural abnormalities, or altered growth in the foetus.
- Period considered: from preimplantation to the day before birth.
- Endpoints: litter composition (e.g. resorptions, live, dead foetuses), embryonic development, foetal growth, morphological variations and malformations. Functional deficits are not considered.
- Species: rodent (preferably rat) and non-rodent (preferably rabbit).
- It is the main guideline used for cosmetic testing of developmental and reproductive toxicity.

OECD test guideline 415: one-generation reproduction toxicity study (OECD 1983)

- Studies the effects on male and female reproductive performance, such as gonadal function, oestrous cycle, mating behaviour, conception, parturition, lactation and weaning. It may also provide preliminary information about developmental toxic effects, such as neonatal morbidity, mortality, behaviour and teratogenesis and to serve as a guide for subsequent tests.
- Period considered: continuously over one generation.
- Endpoints: growth, development and viability; pregnancy length and birth outcome; histopathology of sex organs and target organs; and fertility.
- Preferred species: rat or mouse.
- It is only used in some cases for cosmetic testing.

OECD test guideline 416: two-generation reproduction toxicity (OECD 2001b)

- Studies the effects of a substance on the integrity and performance of the male and female reproductive systems, and on the growth and development of the offspring, including gonadal function, the oestrus cycle, mating behaviour, conception, gestation, parturition,

lactation, and weaning, and the growth and development of the offspring. It may provide information on neonatal morbidity, mortality, and preliminary data on prenatal and postnatal developmental toxicity.

- Period considered: continuously over two or several generations.
- Endpoints: growth, development and viability; pregnancy length and birth outcome; histopathology of sex organs and target organs; fertility; and oestrus cyclicity and sperm quality.
- Preferred species: the rat.
- It is only used in some cases for cosmetic testing.

OECD test guideline 421: reproduction/developmental toxicity screening test (OECD 1995)

- Generates preliminary information concerning the effects of a substance on male and female reproductive performance such as gonadal function, mating behaviour, conception, development of the conceptus and parturition. It is not an alternative to, nor does it replace the Test Guidelines 414, 415 and 416. Positive results are useful for initial hazard assessment and contribute to decisions with respect to the necessity and timing of additional testing.
- Period: from 2 weeks prior to mating until day 4 postnatally.
- Endpoints: fertility; pregnancy length and birth outcome; histopathology of sex organs and target organs; foetal and pup growth and survival until day 3.
- Preferred species: the rat.
- Not commonly used for cosmetics but required under the REACh regulation.

OECD test guideline 422: combined repeated dose toxicity study with the reproduction/developmental toxicity screening test (OECD 1996)

- Apart from gonadal function, mating behaviour, conception, development of the conceptus and parturition, the Guideline also places emphasis on neurological effects.
- Useful as part of the initial screening of chemicals for which little or no toxicological information is available and can serve as an alternative to conducting two separate tests for repeated dose toxicity (TG 407) and reproduction/developmental toxicity (TG 421), respectively. It can also be used as a dose range finding study for more extensive reproduction/developmental studies, or when otherwise considered relevant. It will not provide evidence for definite claims of no reproduction/developmental effects.

- Period: from 2 weeks prior to mating until day 4 postnatally.
- Endpoints: fertility; pregnancy length and birth outcome; histopathology of sex organs and target organs and brain; foetal and pup growth and survival until day 3.
- Preferred species: the rat.
- Not commonly used for cosmetics but required under the REACh regulation.

OECD test guideline 426: developmental neurotoxicity study (OECD 2007c)

- Study the potential functional and morphological effects on the developing nervous system of the offspring of repeated exposure to a substance during in utero and early postnatal development. It can be conducted as a separate study, incorporated into a reproductive toxicity and/or adult neurotoxicity study (e.g. TG 415, 416, 424), or added onto a prenatal developmental toxicity study (e.g. TG 414).
- Period: during pregnancy and lactation.
- Endpoints: pregnancy length and birth outcome; physical and functional maturation; behavioural changes due to CNS and PNS effects; and brain weights and neuropathology.
- Preferred species: the rat.
- It is only regulatory required for the evaluation of agrochemicals and it is unlikely that it will be used for the safety evaluation of cosmetic ingredients.

OECD test guideline 440: uterotrophic bioassay in rodents: a short-term screening test for oestrogenic properties (OECD 2007d)

- This in vivo test evaluates the ability of a chemical to elicit biological endocrine disruption activities consistent with agonists or antagonists of natural oestrogens (e.g. 17 β -estradiol). It is based on the increase in uterine weight or uterotrophic response. The uterus responds to oestrogens with an increase in weight due to water imbibition, followed by a weight gain due to tissue growth.
- Endpoint: uterotrophic response to oestrogens.
- Preferred species: rat (or mature mice).
- It is only used in some cases for cosmetic testing.

OECD test guideline 441: Hershberger bioassay in rats: a short-term screening assay for (anti-) androgenic properties (OECD 2009b)

- This in vivo test evaluates the ability of a chemical to elicit biological endocrine disruption activities

consistent with androgen agonists, antagonists or 5 α -reductase inhibitors.

- Endpoints: changes in weight of five androgen-dependent tissues in the castrate-peripubertal male rat: the ventral prostate, seminal vesicle (plus fluids and coagulating glands), levator ani-bulbocavernosus muscle, paired Cowper's glands and the glans penis.
- Preferred species: the rat.
- It is only used in some cases for cosmetic testing.

OECD test guideline 455: the stably transfected human oestrogen receptor- α transcriptional activation assay for detection of oestrogenic agonist-activity of chemicals (OECD 2009c)

- This in vitro assay evaluates the transcriptional activation mediated by the hER α of oestrogen responsive genes, a process considered to be one of the key mechanisms of possible endocrine disruption related health hazards. The assay provides mechanistic information and can be used for screening and prioritisation purposes of oestrogenic compounds.
- Endpoint: induction of hER α -mediated transactivation of luciferase gene expression.
- Test system: the hER α -HeLa-9903 cell line derived from a human cervical tumour and stably transfected.
- It is only used in some cases for cosmetic testing.

Draft OECD test guideline extended one-generation reproductive toxicity study (OECD 2009d)

- This study may eventually replace the 2-generation study in current testing strategies. It will also result in considerable refinement of the study design through the addition of a series of novel parameters and the assessment of many parameters in more animals per litter than currently prescribed in the 2-generation study.
- This procedure will possibly reduce the number of animals by 40%.

Inventory of alternative methods

In the following, alternative methods designed to detect developmental and reproductive toxicants are described. The chapter is divided into tests for developmental toxicity, for placental toxicity and transport, for preimplantation toxicity including fertility, for effects on the endocrine system and finally in silico methods. The list comprises an inventory of in vitro and in silico methods that are described in the literature and/or that are used by chemical

industry. However, these tests do not necessarily have an application in the safety assessment of cosmetic ingredients. See also Table 11.

Developmental toxicity

Whole embryo tests Whole embryo tests require the use of material from living animals and can in that sense not be regarded as full animal-free alternatives. However, the embryos used for exposures in the tests are not considered as experimental animals under current legislation in view of the early stages of embryogenesis used, at which they are not living independently but are still dependent on either maternal or yolk feeding support. The principal advantage of such assays is in the use of an intact embryo that is exposed in vitro, allowing the study of malformations as they could occur in real life.

The rodent whole embryo culture test: Rodent postimplantation whole embryo culture (WEC) is the only available ex vivo test that covers the critical phase of organogenesis in a complete mammalian embryo. It is widely used both in mechanistic studies and as a screening test for developmental toxicants. Gestation day 10–12 rat embryos are cultured during organogenesis in vitro and treated with test chemicals. Endpoints used in the WEC are a series of well-defined morphological endpoints: all tissues receive a score dependent on their developmental stage, and all scores added up give the so-called Total Morphological Score (TMS). Besides this score, malformations and size measurements are noted, the latter comprising of yolk sac diameter, head length and crown-rump length (Brown and Fabro 1981).

The protocol of the WEC was standardised (Anon 2010b) and scientifically validated according to the EC-VAM validation criteria (Genschow et al. 2002). However, the predictability and applicability domains of the WEC are not sufficiently defined yet to allow regulatory implementation. New developments include transcriptomics analyses to improve predictability and to better define the applicability domain of the WEC (Luijten et al. 2010; Robinson et al. 2010).

The zebrafish embryo teratogenicity assay: The zebrafish (*Danio rerio*) embryo has much promise as an in vitro model to investigate the developmental toxicity potential of substances on the developing vertebrate organism (Nagel 2002). Primary endpoints are lethality, malformations and growth retardation. The development of the zebrafish embryo is very similar to the embryogenesis in higher vertebrates, including humans, and many molecular pathways are evolutionary conserved between zebrafish and humans (Zon and Peterson 2005). This method is used not only as a screening tool for teratogenicity (Brannen et al. 2010; Selderslaghs et al. 2009), but also as a means of

Table 11 Inventory of available alternative methods for reproductive toxicity testing

Current endpoint addressed in the animal test	Alternative tests available	Part of mechanism covered	Area of application	Status ^a	Estimated time to enter prevalidation process ^b	Comments
OECD TG 416, OECD TG 415	Steroidogenesis assay using the human H295R adrenocarcinoma cell line	Detects effects of chemicals on steroidogenesis including the production of testosterone, estradiol	Hazard identification and mechanistic studies	Regulatory acceptance in progress		Limitations due to lack of relevant receptors
OECD TG 416, OECD TG 415	Other steroidogenesis assays	Covering different aspects of the steroid production incl. gonadotrophin-receptor-mediated toxicity	Hazard identification and mechanistic studies	R&D	>5	Different tests are based on various cell lines derived from rat, mouse and human tissues, as well as genetically engineered cells and primary Leydig cells
OECD TG 416, OECD TG 415	Placental aromatase assay	Inhibition of androgens aromatisation	Mechanistic studies	Validation ongoing		Different cellular assays as well as microsomal tests are in development
OECD TG 416, OECD TG 415	Thyroid receptor interactions	Binding assays and reporter gene assays	Mechanistic studies	Optimisation	>5	
OECD TG 416, OECD TG 415	Interaction with the aryl hydrocarbon receptor	Induction/inhibition of reporter gene product	Mechanistic studies	Optimisation	3–5	
OECD TG 414						
OECD TG 416, OECD TG 415	Interaction with oestrogen receptors	Oestrogen receptor binding, induction/inhibition of reporter gene product, cell proliferation, Ishikawa cell test	Mechanistic studies	Validation and regulatory acceptance ongoing		
OECD TG 414						
OECD TG 440						
OECD TG 416, OECD TG 415	Interaction with the Progesterone receptor	Receptor binding, reporter gene assays	Mechanistic studies	Optimisation	3–5	
OECD TG 414						
OECD TG 416, OECD TG 415	Progesterone production	Cell viability, progesterone production, histology	Mechanistic studies	R&D	>5	
OECD TG 414						
OECD TG 416, OECD TG 415	Interaction with androgen receptor	Binding and induction/inhibition of reporter gene product	Mechanistic studies	Validation and regulatory acceptance ongoing		
OECD TG 414						
OECD TG 441						

Table 11 continued

Current endpoint addressed in the animal test	Alternative tests available	Part of mechanism covered	Area of application	Status ^a	Estimated time to enter prevalidation process ^b	Comments
OECD TG 416, OECD TG 415	Gonadotrophin-mediated effects	Release of gonadotrophins or receptor binding	Mechanistic studies	R&D	>5	Several cellular tests based on primary cells/explants or cell lines covering FSH, GnRH, LH, induced effects are described
OECD TG 416, OECD TG 415	Placenta perfusion system	Human ex vivo model to assess transplacental transfer indicating foetal exposure and metabolism of test chemicals	Hazard identification that might lead to fertility impairments	Optimisation	3–5	
OECD TG 414						
OECD TG 416, OECD TG 415	Trophoblast cell assay	Toxicity assessment during embryo implantation		R&D	>5	
OECD TG 414						
OECD TG 416, OECD TG 415	Mouse embryo bioassay	Fertilisation, first cleavage embryo development	Hazard identification that might lead to fertility impairments	Optimisation	3–5	
OECD TG 416, OECD TG 415	Follicle bioassay (FBA)	Completion of oocyte meiosis up to the metaphase II Oogenesis: oocyte yield; diameter; nuclear maturation	Hazard identification that might lead to fertility impairments	Optimisation	3–5	
OECD TG 416, OECD TG 415	Bovine maturation test	Completion of oocyte meiosis up to the metaphase II	Hazard identification that might lead to fertility impairments	Pre-validation	0	
OECD TG 416, OECD TG 415	Bovine fertilisation test	Formation of female and male pronuclei after penetration of capacitated bull spermatozoa into matured oocytes	Hazard identification that might lead to fertility impairments	Optimisation	3–5	
OECD TG 416, OECD TG 415	ReProComet assay	DNA strand breaks and alkali-labile sites in bull sperm	Hazard identification that might lead to fertility impairments	Optimisation	3–5	
OECD TG 416, OECD TG 415	CASA test	Multiple leading to impairments in motility, and viability as well as morphology of spermatozoa	Hazard identification that might lead to fertility impairments	Pre-validation	0	
OECD TG 416, OECD TG 415	Sertoli cell tests	Secretion of Inhibin B by (1) primary rat Sertoli cells and (2) cells of line SerW3; cytotoxicity	Hazard identification that might lead to fertility impairments	R&D	>5	
OECD TG 414	Stem cell-based tests for assessing embryotoxicity	Interaction with cell differentiation into neural, cartilage, bone and cardiac cells, chemical effects on signalling pathways	Hazard identification of embryotoxicants	Optimisation	>3–5	Stem cell-based systems are difficult to optimised due to the challenges in stem cell differentiation protocols

Table 11 continued

Current endpoint addressed in the animal test	Alternative tests available	Part of mechanism covered	Area of application	Status ^a	Estimated time to enter prevalidation process ^b	Comments
OECD TG 414	Murine embryonic stem cell test (EST)	Detects effects on differentiating embryonic stem cells into cardiomyocytes and cytotoxicity	Hazard identification of embryotoxicants	Validated	0	
OECD TG 414	Whole rat embryo culture		Hazard identification of embryotoxicants	Validated		
OECD TG 414	Chicken embryo culture		Hazard identification of embryotoxicants	Pre-validation	0	
OECD TG 414	The Zebrafish Embryo Teratogenicity Assay	Assessment of lethality, malformations and growth retardation	Screening test for the hazard identification of developmental toxicants	Optimisation	>3–5	A common standard protocol is needed to start prevalidation
OECD TG 414	Frog Embryo Teratogenesis Assay Xenopus (FETAX)	Assessment of lethality, malformations and growth retardation	Screening test for the hazard identification of developmental toxicants	Pre-validation	0	
OECD TG 416, OECD TG 415	Hamster Egg Penetration Test/ Hypoosmotic Swelling Test	<i>Functional status of the sperm</i> Sperm quality Fertilising capacity	Hazard identification leading to female and/or male toxicity	Optimisation	3–5	

Definitions used with regard to the development, validation and acceptance process of alternative methods are based on the description of the original document (Zuang and Eskes 2005). In summary a test is still in the R&D phase if basic research and development of a method and/or the test protocol is still ongoing. The phase of optimisation aims to prepare in vitro tests to fulfil ECVAM's criteria for entering into prevalidation. The prevalidation is a small-scale interlaboratory study to assess the readiness of a test for inclusion in a formal large scale validation study (Curren et al. 1995; Hartung et al. 2004). The time to enter the prevalidation process was estimated assuming optimal conditions. Generally high fall-out rates occur during the phases of test development and after entering the prevalidation phase. The list of alternative assays in this table comprises only a small proportion of the tests anticipated to be needed to replace in vivo reproductive toxicity testing. Only tests for which publications are available since 2005 were included. Estimations regarding the status of the different tests and timeframe have been performed by the working group and not officially done by the ECVAM

^a Five levels in the development of a test have been defined: Research and Development (R&D), Test optimisation, prevalidation, validation, regulatory acceptance

^b Under optimal conditions

investigating specific mechanisms related to the teratogenic potential of certain substances.

In principle, the fertilised fish eggs are exposed to different concentrations of a test substance. At different time points, the exposed developing fish embryos are observed and scored for lethal, embryotoxic and/or teratogenic effects. Several protocols have been published differing in e.g. (1) the start and duration of exposure to the test substance, (2) the use of complete or dechorionated fish embryos, (3) the presence or absence of a metabolic activation system (Busquet et al. 2008) or (4) the scoring system and observation intervals.

The zebrafish embryo teratogenicity assay is increasingly used by many laboratories in academia and industry. Intralaboratory studies demonstrate good concordance/

predictivity in correctly classifying in vivo teratogens and non-teratogens (Augustine-Rauch et al. 2010; Selderslaghs et al. 2009). An important step forward would be the agreement on a common standard protocol, which is the prerequisite of a successful prevalidation. Currently, a consortium from pharmaceutical industry has been established to share results and facilitate harmonisation of this promising in vitro method (Augustine 2009).

Frog embryo teratogenesis assay xenopus (FETAX): The FETAX is a whole embryo screening assay, based on the South African clawed frog *Xenopus laevis*, to identify substances that may pose a developmental hazard in humans (Bantle et al. 1999). According to the American Society for Testing and Methods (ASTM) guidelines (Anon 1998), fertilised eggs in the mid- to late-blastula

stage are incubated in media containing the test substance for 96 h. The embryos are scored for lethality, growth retardation and malformations at different time points. Similar to the zebrafish embryo teratogenicity assay, FETAX encompasses organogenesis and does not include later events of development.

In an interlaboratory validation study using 12 compounds, FETAX yielded repeatable and reliable data. However, transferability is still an issue of concern. The inclusion of a mammalian metabolic activation system was essential for the correct prediction of the teratogenic potential of substances. However, FETAX still requires further development (Bantle et al. 1999). Efforts have to be made to improve the predictability of this assay (Fort and Paul 2002).

The chicken embryotoxicity screening test (CHEST): The chicken embryotoxicity screening test (CHEST) was first described in 1976 by Jelinek et al. as a fast and cheap teratogenicity test (Jelinek 1977). In the first protocol described, CHEST comprised two phases of testing, i.e. CHEST I, which determines the toxic dose range in very early administration time (24 h) and CHEST II that determines the teratogenic dose range and covers late effects on the embryo development (days 2, 3 and 4). Recently, adaptations of this protocol were developed (Boehn et al. 2009).

The main endpoints assessed using the modified CHEST are mortality, malformations, embryo development, blood vessel development and blood vessel coloration. Compounds or mixtures can easily be administered to the windowed eggs, and effects on the developing embryo can be investigated. Moreover, the chick embryo possesses its own basic metabolic capacity providing the possibility to screen for metabolites (Kotwani 1998). Studies of Bernshausen et al. (2009) revealed metabolic activities of cytochrome P450 (CYP) and glutathione S-transferases (GST) in 72-h-old chicken embryo sub-cellular fractions.

However, the chick embryo in ovo system has been criticised for not being able to distinguish general toxicity from specific developmental effects and the absence of mammalian maternal–foetal relations (Anon 1967). In addition, CHEST produces a high rate of false positives especially among irritant and corrosive substances that show an evident effect on the blood vessels of the chick embryo (Boehn et al. 2009).

Several studies have evaluated the CHEST and similar protocols (Durmus et al. 2005; Gilani and Alibhai 1990; Jelinek et al. 1985; Jelinek and Marhan 1994; Kemper and Luepke 1986; Kucera and Burnand 1987), and CHEST was demonstrated to be a reproducible test system that delivered quantifiable data for evaluation. At the present time, some laboratories in academia and industry are using

CHEST for routine embryotoxicity screening purposes and mechanistic studies.

The micromass test The micromass test (MM) is making use of cell cultures of the limb bud and/or neuronal cells (Flint 1983; Flint and Orton 1984). The cells are isolated from the limb or the cephalic tissues of mid-organogenesis embryos. After preparing a single cell solution, the cells are seeded in a high density and undergo differentiation into chondrocytes and neurons without additional stimulation. The viability and differentiation after exposure to test chemicals is analysed by measuring the neutral red uptake as well as by staining differentiated cells with alcian blue. The intensity of staining is analysed by using spectrophotometric methods (Brown et al. 1995).

The protocol using micromass cultures of the limb buds has been validated in an ECVAM validation study (Genschow et al. 2002). Data on intra- and interlaboratory variability, transferability and in vivo/in vitro comparisons are available. The number of laboratories currently using the MM is limited.

Pluripotent stem cell-based in vitro tests The potential of embryonic stem cells to differentiate into all cell types of the mammalian organism (pluripotency) provides the scientific rationale to assess adverse effects on the differentiating embryonic stem cells that might be relevant for embryotoxicity in vivo. In 2002, the embryonic stem cell test (EST) that is based on the cytotoxicity assessment as well as the evaluation of differentiation inhibition into cardiomyocytes was scientifically validated (Balls and Hellsten 2002). However, postvalidation evaluations have shown that the applicability domain and its predictive capacity have not yet been sufficiently defined and that the original prediction model has to be modified (Marx-Stoelting et al. 2009). Nevertheless, various industrial sectors are using other methods involving ES cell differentiation for predicting embryotoxicity for pre-screening (Paquette et al. 2008). These embryonic stem cell tests vary in their readouts but also in the target cell differentiation (Peters et al. 2008; Zur Nieden et al. 2004). Depending on the area of application, effects on differentiating neural cells (Stummann et al. 2009b; Theunissen et al. 2010), cardiomyocytes (Buesen et al. 2009) and skeletal cells (Stummann et al. 2009b; Zur Nieden et al. 2004; Zur Nieden et al. 2010) have been investigated. Effects on the quantity of differentiated target cells have been assessed by using immunological methods such as flow cytometry (Buesen et al. 2009) or molecular biological methods such as RT-PCRs and omics (Chapin et al. 2007; Osman et al. 2010; van Dartel et al. 2009; van Dartel et al. 2010; West et al. 2010; Winkler et al. 2009; Zur Nieden et al. 2001; Zur Nieden et al. 2004). Several of the methodologies could

also be automated in order to increase the throughput of substances and make the test available for screening purposes (Peters et al. 2008).

In addition, the establishment of human embryonic stem cell-based tests should contribute to a detailed understanding on mechanisms leading to human developmental toxicity which should substantially contribute to a better hazard identification/characterisation for humans. In 2007, Cezar and colleagues applied hESCs with metabolomics approaches for developmental toxicity testing (Cezar et al. 2007). They were able to identify alterations in the metabolic profile of hESCs exposed to developmental toxicants. This study highlights the possibility of using omics technologies in combination with ESCs and ESC-derived differentiated cells as a novel tool to identify predictive biomarkers for efficacy and safety assessment of substance.

The generation of genetically engineered embryonic stem cell lines allows an easy monitoring of toxic effects in medium-throughput applications. For example, the generation of transgenic cell lines that are using a heart-cell-specific promoter/enhancer controlling the expression of reporter genes allows measuring quantitatively side effects on differentiating heart cells through a reduction in fluorescence (Bremer et al. 1999). Another class of reporter gene assays such as the ReProGlo assay detects chemical-induced alterations in the canonical Wnt/ β -catenin signalling pathway, which is involved in the regulation of early embryonic development (Uibel et al. 2009). The development of additional genetically engineered embryonic stem cell lines evaluating biologically significant perturbations in key toxicity pathways of embryotoxicity might follow and will provide a mechanistic understanding on developmental toxicity. Nevertheless, also these tests are still in their research and developmental phase.

The establishment of stable differentiation protocols is challenging and requires additional scientific work. Considerable scientific/technical efforts are currently ongoing to stabilise stem cell differentiation mainly for application in regenerative medicine. Due to the growing knowledge in stem cell technologies, progress can be expected in the next couple of years. First indications that successful tests can be developed have been published (Adler et al. 2008a, 2008b; Stummann and Bremer 2008; West et al. 2010).

Placental toxicity and transport

The placental perfusion assay Understanding the placental transport of compounds provided to the pregnant mother is essential to reduce the risks of foetal exposure to harmful substances during pregnancy. The placenta serves as the interface between the maternal and foetal circulations during pregnancy. Ex vivo human placental perfusion provides an opportunity to carry out research without

ethical difficulties. It takes around 30 min following the birth to set up a perfusion, and the perfusion conditions allow for continued placental tissue viability for several hours. Viability of the placenta during the experiments is verified by monitoring leakage from the foetal compartment, oxygen transfer and glucose consumption. Appropriate antipyrine transfer between the maternal and foetal circulations confirms proper experimental set up and can be used to normalise differences between placentas. Other advantages of placental perfusion experiments include the retention of in vivo placental organisation and assessment of binding to placental tissue (Mose et al. 2008; Myllynen et al. 2009). However, the application of this assay is limited due to placenta to placenta variations and the limited relevance of the term placenta for the period of embryonic development. Due to the complexity of the assay, it is not applicable to routine testing of high numbers of test compounds.

Trophoblast cell assay In this assay, the BeWo cell line is used which represents an immortalised trophoblastic line of human origin. The cells form polarised, confluent monolayers and have proven useful in transport studies. The assay based on BeWo cells serves as an in vitro model of the rate-limiting barrier to maternal–foetal exchange. The BeWo b30 model consists predominantly of cytotrophoblast cells which form a confluent monolayer with tight junctions, but they do not spontaneously differentiate to syncytiotrophoblasts, and the model lacks the connective tissue which is present in vivo (Morck et al. 2010).

Preimplantation toxicity

Male fertility Computer-assisted sperm analysis: The computer-assisted sperm analysis (CASA) allows to monitor effects of chemicals on spermatozoa with possible implications on fertility. Potential viability, motility, velocity, motion, and morphology of mammalian semen will be analysed in real time. This allows the detection of reversible and irreversible damages (recovery effect) to the mature sperm as well as repeated dose effects. For reproductive medicine, fully automated semen analysers are available. Several chemicals have already been tested in different laboratories, and an INVITTOX protocol is available. The test has been evaluated by two independent laboratories by testing more than 35 test chemicals (Anon 2010a). The lower sensitivity of mature sperm in comparison with earlier stages of spermatogenesis must be considered and may limit the relevance of this test.

Leydig cell assay: A disturbance of the endocrine system due to effects of chemicals on steroidogenesis or due to specific cytotoxic effects on Leydig cells leads to a decreased development of spermatozoa and impaired

fertility since Leydig cells nurture the developing sperm cell. A new Leydig cell line, BLT1-L17, that responds very well and quite robustly to luteinising hormone (LH) or its analogue, chorionic gonadotropin (hCG), has been characterised.

In the assay, the MTT test serves as a general toxicity endpoint and testosterone production as the Leydig cell-specific endpoint. BLT1-L17 cells were exposed to 15 chemicals, and the data obtained with this set of test chemicals indicate that the cell line is a candidate for further development into a rigorous test applicable for in vitro reproductive toxicity assessment acting via interference with testosterone production (Anon 2010a). Another Leydig cell system has also been developed and proven to be applicable to the analysis of oestrogenic agents (La et al. 2010).

Sertoli cell assay: Sertoli cells form the basis of the blood–testis barrier and divide the tubular area into adluminal and basal compartments protecting the maturing germ cells from chemical insults. In the assay, rat primary cultures and the SerW3 line are used. The Sertoli cell assay was developed by pharmaceutical industry and transferred to a second laboratory. General cytotoxicity and the secretion of inhibin B are measured. These two endpoints allow a classification of test chemicals as positive or negative for testicular toxicity. In addition, the integrity of tight junctions forming the blood–testis barrier can be studied in the SerW3 cell line, providing a new endpoint to study the mechanism of action of testicular toxicants. Further studies are needed to fully understand the utility of this test (Anon 2010a).

Very recently, a new 3D-culture system has been developed composed of both Sertoli and germ cells, which may allow mimicking to some extent the in vivo blood–testis barrier (Legendre et al. 2010).

ReProComet assay: The ReProComet assay (Repair Proficient Comet assay) was developed to detect chemically induced DNA damage in sperm cells. In order to circumvent the intrinsic repair deficiency of the sperm cells, a strategy is deployed involving a supplementation with protein extract from somatic cells after the chemical treatment. Liquid nitrogen frozen bull sperm is used for the analysis. Bull sperm is incubated with the test chemicals for 2 h. A SYBR-14/Propidium iodide flow cytometric analysis is used to evaluate sperm viability in addition to the four comet assay endpoints tail length, tail moment, fraction of tail DNA, fraction of head DNA (Anon 2010a; Cordelli et al. 2007). The rationale of the test design needs further clarification.

Female fertility Follicle culture bioassay (FBA): The FBA allows multiparametric in vitro analysis of effects of chemicals on the ovarian function such as folliculogenesis,

steroidogenesis and oogenesis. Mouse ovarian pre-antral follicles are grown in vitro until the preovulatory stage followed by in vitro ovulation induction and mature oocyte retrieval. During the in vitro growth period (12 days), the follicles develop with theca cell proliferation, granulosa cell proliferation and differentiation, meanwhile supporting oocyte growth and maturation. In the FBA, the in vitro growing follicles are exposed to chemicals in a chronically or acute manner, and effects on the different biological processes of folliculogenesis, steroidogenesis and oogenesis are analysed with morphological, biochemical and functional parameters. The FBA is still in the phase of development. It requires further standardisation and transferability to other laboratories has to be addressed (Anon 2010a; Lemeire et al. 2007).

In vitro bovine oocyte maturation assay (bIVM): The bIVM assay focuses on the use of bovine oocytes for toxicity testing during the process of oocyte maturation in vitro. The test screens for potential adverse effects on the process of oocyte maturation after exposure of cumulus–oocyte complexes to test substances, with special reference to nuclear configuration changes within the oocyte as compared to control non-exposed oocytes. Endpoint is the successful achievement of the maturation stage metaphase II (completion of meiosis up to the metaphase II). The interlaboratory variability and the transferability of the bIVM test was analysed for a set of eight chemicals, and the statistical analysis of the data obtained from the two laboratories demonstrated that there was a good concordance of results across the laboratories (Anon 2010a; Lazzari et al. 2008; Luciano et al. 2010). Testing of additional compounds is necessary in order to assess the predictivity of this test.

In vitro bovine fertilisation test (bIVF): The bIVF assay focuses on the use of bovine oocytes and sperms for toxicity testing during the process of in vitro fertilisation. The purpose of the test is to (1) screen for adverse effects of chemicals on the process of oocyte fertilisation and (2) investigate the mechanism of action of reproductive toxicants. Both oocytes and sperms are exposed to test chemicals; therefore, the adverse effects on the function of both gametes can be monitored. Specific endpoints are (1) penetration of capacitated bull spermatozoa into matured oocytes and (2) formation of the female and male pronuclei (Lazzari et al. 2008). This test is still in a very early phase of development and further investigations are necessary to assess its toxicological relevance (Anon 2010a).

Mouse peri-implantation assay (MEPA): The mouse peri-implantation assay is an in vitro bioassay that allows studying the effect of compounds on the development of the pre-implantation embryo and its capacity to survive upon hatching around the implantation period. The assay is based on the in vitro culture of mouse zygotes. The zygotes

are cultured in groups of 10 for 7 days with daily observation and scoring of embryo development. These daily morphological observations allow pinpointing potential deviations of the timely regulated pre-implantation embryo. The bioassay has a high intra-laboratory reproducibility. It allows the characterisation of the sensitive stage of embryo development (Van Merris et al. 2007). The MEPA is still in the phase of development. It requires further standardisation and transferability to other laboratories has to be addressed (Anon 2010a).

In vitro tests for assessing effects on the endocrine system

Ishikawa cell test The human endometrium is a fertility-determining factor. Its receptivity during the implantation window may be altered by chemicals. The Ishikawa cell test aims to identify chemicals which alter the expression of embryo-implantation-associated target genes in human endometrial adenocarcinoma Ishikawa cells. Ishikawa cells are cultured to subconfluency and incubated for 0.5–24 h with test substances. This test system is a tissue-specific model to detect oestrogenic activity of chemicals which upregulate progesterone receptor (PR) mRNA in the human endometrium. The Ishikawa model is informative regarding the mode of action of positive tested chemicals and provides guidance for prioritisation for further testing (Schaefer et al. 2010). The Ishikawa cell test is still in the phase of development. It requires further standardisation and transferability to other laboratories has to be addressed.

Cell proliferation based assays for testing oestrogen activity Oestrogenic activity of substances can be assessed by measuring in vitro proliferation of cell lines containing the ER- α and ER- β oestrogen receptors such as the human breast cancer cell line MCF-7. The binding of the natural hormone or other oestrogen like xenobiotics leads to conformational changes that allow the oestrogen-ligand complex to proceed from inactive proteins to active transcriptional regulators that induce transcription of oestrogen responsive genes which lead to an oestrogen-dependent proliferation of cells (Lippman et al. 1976). One example of these assays, using MCF-7 cells, is currently undergoing a validation study coordinated by NICEATM.

Receptor binding assays Relevant hormonal receptors can be isolated either from primary tissues such as rat prostate (US EPA 2007a) or generated with recombinant technologies (Hartig et al. 2008). Nevertheless, all tests rely on the same principles assessing the competitive binding of a substance to a receptor of interest.

Most advanced are receptor binding tests based on the oestrogen receptor. Chemical interactions with the oestrogen receptor might affect the development of female

secondary sexual characteristics and/or the regulation of the menstrual cycle. Several tests such as the uterine cytosol (ER-Rat Uterine Cytosol) assay (US EPA 2009) or the human recombinant full-length oestrogen receptor-alpha binding assay (Freyberger et al. 2010) have been intensively evaluated in (pre)validation trials under the lead of the US-EPA. The regulatory acceptance of oestrogen receptor binding tests is in preparation.

Another important receptor for the endocrine system is the androgen receptor. Androgens are mainly concerned in the development and maintenance of male secondary sexual characteristics. Several receptor binding tests based on isolated proteins from the cytosol of the rat prostate (Battelle 2002) or recombinant proteins (Freyberger et al. 2009a; Hartig et al. 2008) have been developed and optimised. The validation of other androgen receptor binding tests has been taken up in the work programme of the OECD (<http://www.oecd.org/dataoecd/54/29/46034089.pdf>).

Another highly relevant receptor in the context of receptor-mediated reproductive toxicity is the progesterone receptor. As for the previous receptors also for the progesterone, receptor binding assays have been developed in order to assess effects that might have influence on the menstrual cycle, the pregnancy and/or embryogenesis. Currently, several developed assays are available that are using for example rabbit uterine cytosol (Attardi et al. 2006), recombinant receptor (Viswanath et al. 2008) or even whole cells (Klotz et al. 1997).

The thyroid hormone receptor is highly relevant for the development of the central nervous system. Tests monitoring the binding of the thyroid hormone triiodothyronine (T3) to its receptor are in the development phase, using recombinant proteins (Ishihara et al. 2009).

Other hormonal receptors playing a key role are binding hormones produced by the hypothalamus (gonadotropin-releasing hormone) or pituitary gland (follicle-stimulating hormone, luteinising hormone). These tests are still in the phase of research and development but need to be considered since these hormones are involved in the feedback loop controlling the reproductive system. Even if biologically highly relevant, further assessments are needed to clarify if they act as major targets for xenobiotics (toxicological relevance).

Transcriptional tests In contrast to the receptor binding tests which only provide information on the binding capacity of a substance to a particular hormone receptor, the so-called transcriptional activation assays are able to distinguish between agonist and antagonistic effects of xenobiotics. The basic principle of transcriptional assays relies on genetically engineered cells which express hormone receptors as well as reporter genes driven by

hormone responsive genes. The intensity of the receptor binding can be measured for example by using spectrophotometric techniques.

This basic principle has been used for the development of several transcriptional tests involving various hormones which are in different stages of standardisation and validation. The oestrogen receptor (ER) transcriptional assays for example quantify the induction of a reporter gene product by the test substance or reference oestrogen. The antagonism is measured by the inhibition of the reference oestrogen induction of the reporter gene or cell proliferation.

Most advanced in the class of transcriptional assays is for example the “LUMI-CELL” test that is currently undergoing a formal validation study by ICCVAM (http://www.iccvam.niehs.nih.gov/methods/endocrine/end_eval.htm). The process of regulatory acceptance of this test is already included in the OECD work plan 2009 of the Test Guidelines Programme. Other tests that will certainly also contribute to a performance-based test guideline are tests named “MELN” (Witters et al. 2010) and “ER α CALUX” (van der Burg et al. 2010b). Anti-oestrogenic activities can also be mediated through the activation of the aryl hydrocarbon receptor. Transcriptional activation assays of this receptor are in optimisation phase, using different cell lines (Bittner et al. 2009; Long et al. 2003).

Similar to ER transcriptional assays, androgen receptor transcriptional assays have been designed (Freyberger et al. 2009b; van der Burg et al. 2010a). A Japanese stably transfected transcriptional activation (STTA) assay for the detection of androgenic and anti-androgenic activity of chemicals is under consideration by OECD (included in work plan 2009 of the Test Guidelines Programme). Other transcriptional assays following the same scientific principle are, tests, e.g. assessing the progesterone transcriptional activity (Molina-Molina et al. 2006; Willemsen et al. 2004) or the interaction with the thyroid receptor (Ghisari and Bonfeld-Jorgensen 2009; Shen et al. 2009). These tests are in their early phase of development, and additional work is necessary to optimise the tests.

Tests assessing steroidogenesis In the past years, significant progress has been made by developing in vitro cell-based assays aiming to detect substances that affect the synthesis of the sex steroid hormones. These tests are at different stages in their development. Nevertheless, all tests are designed to identify xenobiotics that have as their target sites components that perturb biochemical pathways. The complexity of potential target enzyme is very high (as shown in the gonadal steroidogenesis pathway; <http://www.kcampbell.bio.umb.edu/lectureI.htm>).

Furthermore, several receptors regulating steroidogenesis are involved and need to be considered as possible

target for endocrine effects (GnRH, LH, and FSH Receptors). Different assays measuring the gonadotrophin-stimulated steroidogenesis are under development, e.g. FSH (Zachow and Uzumcu 2006) or LH (Lambrot et al. 2009).

A cell-based assay on steroidogenesis, using the H295R cells, designed to measure effects on estradiol and testosterone production has been validated and a draft test guideline is currently under discussion (OECD 2010e). Other tests focusing on the aromatase enzyme (CYP19) are under development, e.g. using human placental microsomes (Anon 2007d).

Application of in silico techniques to reproductive toxicology

Existing data There are a number of international efforts to bring together existing toxicological information on reproductive toxicology in an electronic format. For data that are publicly available (i.e. not Confidential Business Information), they may be released via the internet. Therefore, a number of searchable resources have been developed. These resources can be used in at least two ways: to provide existing information on a substance such that testing may not be required and as a source of data for further in silico modelling.

There are a number of important issues when developing and using toxicological databases. The first is ensuring the quality of the information within the database. There are different issues to determination of data quality. The chemical structure and its identifiers (e.g. name, CAS, 2-D or 3-D structure) must be consistent and correct. The structure and ontology of the database must be sophisticated enough to capture the required information regarding a toxicological test, i.e. species, test, duration, dose, purity, effects, etc. In addition, the transfer of information, e.g. from the open literature requires checking and quality assurance. Finally, there is the issue of the quality of the individual data. This last consideration of assigning data quality is a process that may be undertaken by the database user.

There are a number of (meta-) databases that can be searched for toxicological information (including reproductive effects) on single chemicals. Improvement of the quality of existing databases took place during the past few years and meta-databases offer now the possibility to search numerous data resources and compile this information. Of particular note are

- United States Environmental Protection Agency (US EPA) Aggregated Computational Toxicological Resource (ACToR) available from <http://www.actor.epa.gov/actor/faces/ACToRHome.jsp>.

- Organisation for Economic Cooperation Development (OECD) eChemPortal available from <http://www.webnet3.oecd.org/echemportal/>.
- United States Environmental Protection Agency (US EPA) TOXREF: EPA Toxicology Reference Database in Support of ToxCast™ Program available from http://www.epa.gov/comptox/dsstox/sdf_toxref.html.
- ToxCast™ Program available from <http://www.epa.gov/ncct/toxcast/>.
- The US National Institute of Health National Toxicology Program includes reproductive toxicity tests. Study reports can be searched for and downloaded at <http://www.ntp.niehs.nih.gov/>.

There are a number of other databases and toxicological resources. For the development of (Q)SARs for reproductive toxicity, these have, historically, been relatively limited in terms of size and chemical diversity. Examples of databases containing reproductive toxicity data include those developed by Leadscope Inc (<http://www.leadscope.com>) in cooperation with the United States Food and Drug Administration (US FDA). A specific area where more data are available is for endocrine disruption and the binding of chemicals to specific receptors such as the oestrogen and androgen receptor (Xu et al. 2010).

Grouping/category formation One of the simplest predictive in silico approaches is the development of rational groupings of compounds. If the groups can be populated with reliable data, then it may be possible to make interpolations of activity to fill data gaps (Fabjan et al. 2006; Hewitt et al. 2010). The process of forming a grouping (or category) is often termed “category formation” and that of interpolation is termed “read-across”. The key to success for this process is the development of a reliable grouping. For reproductive toxicology, it is recommended that these groupings are based on mechanisms of action should they be known or implied. There are at least two methods to develop groupings for read-across.

Structural analogues: Compounds can be grouped together in terms of structural features, e.g. the presence of a group that is known to elicit a particular reproductive effect. Tools such as the freely available OECD (Q)SAR Application Toolbox (downloadable from: http://www.oecd.org/document/23/0,3343,en_2649_34377_33957015_1_1_1_37465,00.html) are capable of forming such groupings. The OECD Toolbox does not, at this time, contain a “profiler” for reproductive effects, but it does include the capability to group compounds according to individual, or combinations, of structural features. In addition, the OECD Toolbox provides a profiler for oestrogen receptor binding.

Version 2.0 and Version 3.0 of the Toolbox are planned for releases in October of 2010 and 2012, respectively. Key changes scheduled for Version 2 are improved operation of the Toolbox, expanding and refining key mechanistic profilers, and adding new databases (including ones covering reproductive toxicity endpoints). Key changes expected in Version 3 are new functionalities for handling chemical speciation, metabolism, mixtures and the use of 3D-descriptors, profilers for toxicological categories based on symptom data for chronic health effects (including reproductive toxicity) and expanding the (Q)SAR models inventory.

3-D structure similarity: Various methods are available to assess gross molecular “similarity” and provide a quantitative measure. These have been coded into tools such as the freely available ToxMatch software (downloadable from: <http://www.ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=TOXMATCH>). This provides a method to group compounds together when a mechanistic basis may not be available or immediately transparent. Enoch et al. (2009) illustrated how categories could be formed for the prediction of teratogenicity using these techniques.

Structure–activity relationships (SARs) SARs and structural alerts are appropriate to describe fragments of a molecule which are related to a particular effect, e.g. a reproductive toxicity endpoint. These can be utilised to identify a potential hazard in a chemical. Currently, there are few reliable structural alerts developed for reproductive endpoints. Of the more developed, the Derek for Windows/Derek Nexus software from Lhasa Ltd (<https://www.lhasalimited.org/>) currently contains approximately 20 structural alerts for the reproductive toxicity “super-endpoint” (Hewitt et al. 2010). The current number of structural alerts for reproductive toxicity represents only a small proportion of the probable mechanisms of action. While there are currently few structural alerts, they are accurately defined and (in Derek for Windows/Derek Nexus) very well supported by literature and examples of compounds. Therefore, presence of an alert in a molecule should be considered a credible reason for concern; the absence of an alert provides less confidence that a molecule has no hazard associated with it.

QSARs QSARs (and other in silico methods) for predicting reproductive toxicity were reviewed by Cronin and Worth (Cronin and Worth 2008). QSARs for endpoints within reproductive toxicology have been developed by a wide variety of approaches ranging from regression analysis to multivariate analyses. These techniques have been used to make qualitative predictions (i.e. presence or absence of an effect) as well predictions of potency. A broad range of reproductive effects has been studied by

QSAR, in addition to wide areas of chemical space. Generally, QSARs will work best when the chemical space (and by analogy the mechanistic space) is restricted.

Some QSARs form the basis of some expert systems for predicting reproductive toxicity. These include commercial systems such as TopKat and MultiCASE and freely available systems such as CAESAR. The expert systems for reproductive toxicity were reviewed by Cronin and Worth (Cronin and Worth 2008). Expert systems could be used to prioritise compounds for testing—a good illustration for reproductive toxicology is provided by Jensen et al. (2008). Many expert systems make predictions based on large and chemically and mechanistically heterogeneous datasets. While this makes these global models broadly applicable, it makes the assessment of transparency and mechanistic relevant more complex. There may be increased certainty in using various predictions for reproductive toxicity and forming a consensus through a weight of evidence approach (Hewitt et al. 2010).

QSARs for ADME relating to reproductive toxicity: There are numerous QSAR approaches to predict the absorption, distribution, metabolism and excretion (ADME) properties of compounds. The QSAR approaches to predict ADME are mainly developed for the development of pharmaceuticals, these are well summarised by Madden (Madden 2010). A number of these could be applicable to determine the likelihood of significant bioavailability of a toxicant. For instance, much has been written on the prediction of whether a drug will be soluble and/or bioavailable after an oral dose. Simple, freely available, computational screens based on trivial molecular properties, such as the Lipinski rule of 5, can provide an assessment of whether a compound may reach therapeutic or toxic levels. There are a number of metabolic simulators that can predict potential metabolites, e.g. for liver and skin, although none have been yet been developed for placental metabolism.

Specific QSARs have been developed for the transfer (by passive diffusion) of chemicals across the placenta (Hewitt et al. 2007b). Only very limited models are available for blood–testis transfer and other relevant barriers (Cronin and Hewitt 2007). While the accuracy of such models is limited and they should be applied cautiously, these approaches may provide a basis for prioritisation of compounds within chemical groupings or categories (Hewitt et al. 2010).

While these approaches have been developed for pharmaceutical compounds, they could be applied in the context of cosmetic ingredients. This may require some investigations of the relative chemical space to determine the applicability of these approaches.

In silico approaches for endocrine mechanisms of action
Within in silico approaches for reproductive toxicology,

those for endocrine mechanisms of action are relatively best developed (Cronin and Worth 2008; Devillers 2009). Specifically, (Q)SAR have been well developed for the ability of compounds to bind to the oestrogen receptor and also (albeit to a lesser extent) the thyroid and androgen receptors. These techniques range from simple screens that define the structural requirements for binding, through to QSARs for potency and 3-D QSAR models to incorporate receptor binding explicitly. The range of approaches has the capability to predict receptor affinity of compounds.

An advance from traditional QSAR is provided by VirtualToxLab as described by Vedani and Smiesko (Vedani and Smiesko 2009). This is a computationally intensive approach that attempts to bind a target molecule to each of a number of target receptors (e.g. androgen, aryl hydrocarbon, oestrogen α and β , etc.) associated with reproductive hazard as a basis for potency ranking. The individual models are shown to have good statistical fit for training and test sets and provide an estimation of binding affinities for the receptors. They cover different numbers of chemical classes (from two to eighteen) depending on the nature of the data sets. Thus, if deemed to be a valid prediction, the information provided is analogous to an in vitro binding assay. This advances the use of in silico techniques as it provides the possibility of determining a mechanistically important event, e.g. receptor binding. Forming a consensus from individual predictions will increase the applicability to make assessments for the endpoint of reproductive toxicology. Such a consensus relies of the user having confidence that all relevant receptors have been described and modelled. While VirtualToxLab does not state this, the predicted “toxic potential” provides a good estimate of hazard from the data provided.

Current status of in silico approaches for predicting reproductive toxicity In silico approaches to predict reproductive toxicity vary in type, complexity, ease of use and acceptance. It is widely acknowledged that reproductive effects are among the most difficult endpoints to predict in silico. Reasons for this difficulty include a limited number of data to model, lack of knowledge of mechanisms of action, oversimplification of complex outcomes that may have been brought about by different mechanisms or effects. The problems in modelling have been reflected in poor performance of the models in external assessments. For instance, Pedersen et al. (2003) and Hewitt et al. (2010) have demonstrated the poor predictivity of a number of in silico approaches. However, these studies were seldom accompanied by an assessment of whether a molecule was within the applicability domain, etc. and how much confidence could be associated with a particular prediction.

The current status of in silico methods for reproductive toxicity reflects the problems with modelling this endpoint.

Grouping approaches offer a possibility to build local (Q)SARs and perform read-across. These are currently limited and can be performed only for obvious (structural) analogues for which a mechanistic basis may be known or implied. Such grouping approaches could be improved by the better development of profilers for reproductive toxicity. In addition, new approaches may be required to group compounds more successfully according to receptor-mediated effects: these will also need to be able to manage the problems brought about by subtle effects in molecular structure such as enantiomers. Similarly, structural alerts and SARs should be developed on a mechanistic basis to better reflect the number and complexity of toxicological initiating events.

QSARs are available for reproductive toxicity effects although they will probably work best within small groups of compounds which share a common mechanism of action. Other QSARs are available for ADME properties although these require further refinement. There are no predictive models for the metabolic activity of the placenta and other relevant tissues. 3-D QSARs and those for receptor binding affinity are well developed but are seldom able to differentiate between agonistic and antagonistic effects.

In summary, *in silico* approaches aiming to derive simple statistical relationships between complex effects and structural properties are ambitious and may not account for the subtlety in the mechanisms, such as time dependence and receptor binding effects. To provide relevant solutions to predict reproductive toxicity, further alerts and grouping strategies should be developed around the sensitive endpoints identified through the analyses of databases of *in vivo* responses. These could provide information on the key molecular features that are involved across reproductive toxicology to assist in the development of structural alerts, local (Q)SARs and possibly more global QSAR models.

Identified areas with no alternative methods available and related scientific/technical difficulties

Approaches for alternative testing

A significant number of alternative assays have been developed as described in chapter 5. However, their implementation in regulatory toxicity testing has not yet been achieved. As stated in the introduction to this chapter, the reproductive cycle combines a highly diverse multitude of biological processes and mechanisms, each of which has their own time-related sensitivity to xenobiotic exposures. It is therefore a significant challenge to mimic all aspects of the reproductive cycle with *in vitro* and *in silico* assays, which may be considered necessary in order that

reproductive toxicity can be predicted reliably on the basis of alternative assays alone. The classical aim of “one-to-one” replacement of *in vivo* protocols by alternative tests is clearly not feasible for the complex reproductive and developmental toxicity animal study protocols. Alternative approaches are required in which a limited array of most sensitive endpoints are reproduced by a set of alternative assays which, in combination, could provide sufficient background for hazard identification and risk assessment. Other endpoints might be identified for which alternative methods do not yet exist. These endpoints may include among others ADME, spermatogenesis, sperm maturation, HPG axis and maternally mediated effects.

General limitations of in vitro methods for reproductive toxicity testing

It is generally recognised that *in vitro* methods represent only a very simplified picture of reality, i.e. of living organisms. In the case of reproductive toxicity, each *in vitro* model encompasses only a small part of the complex reproductive cycle and not all steps of reproduction are covered by those assays so far. Moreover, many *in vitro* assays, e.g. the receptor binding assays, investigate rather a cellular mechanism which finally might or might not result in an adverse effect *in vivo*. Most *in vitro* assays do not consider the various aspects of absorption, distribution, metabolism and excretion (ADME), which have a tremendous impact on the toxicological profile of a substance. In addition, the distinction between general toxicity and specific developmental effects is difficult to evaluate even with more complex *in vitro* studies such as the whole embryo assays. In general, the influence of the maternal organism including maternal toxicity is not covered by *in vitro* studies. Finally, technical difficulties may occur such as low water solubility of test materials, which have to be considered in the design of each individual *in vitro* assay.

The testing strategy as the future driving force

Classically, many *in vitro* alternatives have been developed based on relatively simple endpoints that were deemed representative of the wider context of for example embryotoxicity. Intrinsically reductionistic assays such as rodent whole embryo culture (WEC), the embryonic stem cell test (EST) and the limb bud micromass (MM) were validated for their prediction of the entire embryotoxicity endpoint. This approach presumes that the endpoints represented in these tests are actually among the critical ones for embryotoxicity prediction in general. Recent experience with the EST has shown that this is an oversimplification, leading to false positives and false negatives and thus limited predictability, showing that the applicability

domain of the assay was more limited than anticipated (Marx-Stoelting et al. 2009). This has led to the insight that rather than focussing exclusively on individual assays and their relevance, it might be more productive, in view of the regulatory implementation of alternative assays, to start from their anticipated role in testing strategies for chemical risk assessment.

Retrospective analyses to select critical endpoints

A wealth of in vivo data has been collected over the past 30 years since the introduction of OECD test guidelines for reproductive and developmental toxicity testing. Databases collecting past experience are being built and allow detailed analysis to assess the relative sensitivity of endpoints in existing animal protocols (Martin et al. 2009). The combination of the most sensitive endpoints should be able to detect nearly all reproductive and developmental toxicants. Current thinking evolves towards designing alternative assays limited to covering each of these most sensitive endpoints only. In addition, these assays should be used appropriately within the entire testing strategy. This can be either tiered (the next test carried out dependent on the outcome of the former), or as a battery (several tests in parallel), or in some combination of the two. They should be performed in a phase within the strategy where the information can be used optimally for hazard and risk assessment in order to preclude as much as possible any ultimate animal experimentation (Bremer et al. 2007).

Towards the definition of novel testing paradigms

The OECD conceptual framework for reproductive toxicity testing provides an outline of such an approach. Starting from in silico (non-testing) information such as physico-chemical characteristics, structure–activity relationships and read-across methods and followed by in vitro assays, it should be possible to restrict in vivo testing to an essential minimum. Information from innovative molecular approaches including omics could enhance the testing strategy. Crucial steps in the process of designing the testing strategy on the basis of alternative approaches are (1) the identification of most sensitive endpoints, (2) evaluation of existing alternative tests containing these endpoints, (3) design of novel assays for sensitive end-

points not already addressed by current assays and (4) optimisation of the testing strategy through the combination of assays. A first example of such an approach is the ReProTect feasibility study (Schenk et al. 2010). It should be realised that the entire reproductive cycle in a living animal is, by definition, more than the sum of any combination of alternative approaches. Therefore, for reproductive toxicity testing, animal studies will remain the last resort for foreseeable time. This is caused by for example the often long lag time between exposure and observed adverse effect in the reproductive cycle and also by kinetic aspects, which hamper ready translation from in vitro effective concentrations to in vivo effective doses. In addition, efforts to build risk assessment solely on human-derived data (Krewski et al. 2010) are currently limited by the scarcity of relevant human toxicological data.

Time schedule for phasing out in vivo reproductive toxicity testing

Given current knowledge and a realistic outlook into the future, full replacement of animal studies for reproductive toxicity hazard assessment is not probable within foreseeable time (>10 years). However, alternative assays are already being used for priority setting and screening purposes. In addition, alternative assays can make an important contribution to the mechanistic understanding of reproductive toxicity. Such tests may actually be able to give more specific information on the interference of the test compound with the endpoint involved than the in vivo study is able to generate. The challenge is to build on these advantages of alternative tests in generating a testing strategy in which the most sensitive endpoints are combined in a well-informed selection of alternative assays (Schenk et al. 2010). Applying such a strategy in a tiered screening situation (cf. the OECD conceptual framework) could preclude the in vivo testing of many reproductive toxicants and thus would considerably refine and reduce testing for reproductive toxicity.

Annex 1

See Table 12.

Table 12 Final list of invited experts to the working groups on alternatives to animal testing/cosmetics—ISPRA

Name	Proposed by
<i>Working group toxicokinetics</i>	
Ulrike Bernauer	EFfCI ^a
George Loizou	Colipa
Emanuela Testai	SCCS ^b
Ursula Gundert-Remy	SCCS
Jos Bessems	MS ^c (NL)
Frederic Y. Bois	JRC ^d
Olavi Pelkonen	MS (FI)
Ester Brandon	JRC
<i>Working group skin sensitisation</i>	
Reinhard Kreiling	EFfCI
Gavin Maxwell	Colipa
Ian Kimber	SCCS
Klaus Ejner Andersen	SCCS
Henk van Loveren	MS (NL)
Aynur Aptula	JRC
Hanna Tähti	MS (FI)
David Basketter	ECEAE ^e
<i>Working group repeated dose—including chronic, sub-chronic and sub-acute exposure</i>	
Vera Rogiers	SCCA, Colipa, EFfCI
Alan Boobis	Colipa
Jos Kleinjans	SCCS
Wolfgang Dekant	SCCS
Greet Schoeters	SCCS
Paolo Mazzatorta	JRC
Hannu Komulainen	MS (FI)
Stuart Creton	ECEAE
Mathieu Vinken	JRC
<i>Working group carcinogenicity</i>	
Susan Felter	Colipa
Joost van Delft	SCCS
Emilio Benfenati	JRC
Tuula Heinonen	MS (FI)
Jan van Benthem	MS (NL)
Albrecht Poth	ECEAE
Stefan Pfuhrer	JRC
<i>Working group reproductive toxicity</i>	
Thomas Broschard	EFfCI
George Daston	Colipa
Aldert Piersma	ECEAE, MS, Colipa
Guillermo Repetto	MS (ES)
Mark Cronin	JRC
Michael Schwarz	JRC
Sarah Adler	MS (DE)

^a EFfCI European Federation for Cosmetic Ingredients

^b SCCS Scientific Committee for Consumer Safety

^c MS Member State

^d JRC Joint Research Centre

^e ECEAE European Coalition to End Animal Experiments

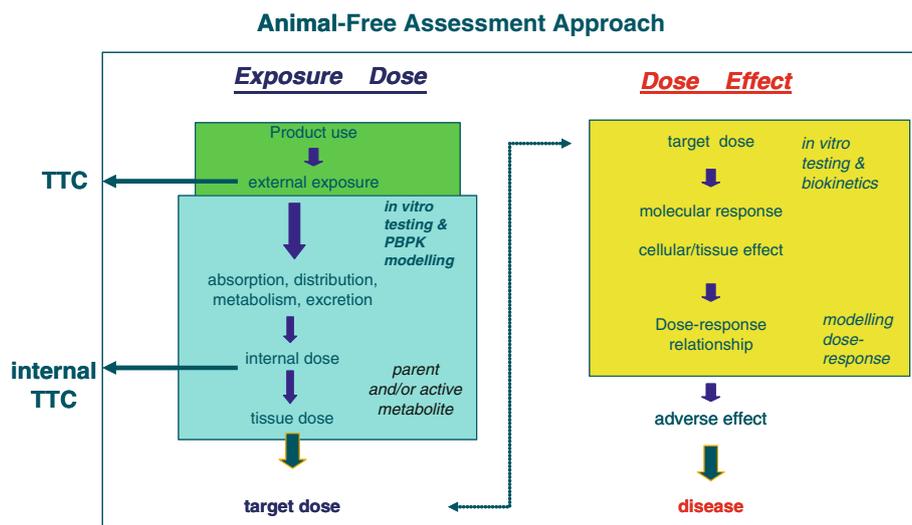


Fig. 11 An animal-free assessment approach that appropriately takes account of toxicokinetics. The *left column* starts on *top* with (external) exposure and ends at the *bottom* with (internal) target tissue dose/concentration. The *right column* starts with the estimated/predicted target tissue dose/concentration which would indicate the range of concentrations that should be tested *in vitro*. If *in vitro*

effects are not measurable at target tissue concentration, any *in vivo* effects and consequently possible health risks would be unlikely. If certain effects are discovered at or above a defined target dose, the *left column* could be applied in reverse order to estimate the external exposure that would be needed to reach that target dose. Adapted from Bessems (2009)

Annex 2

A framework for risk assessment without animal testing

The working group experts underlined in their discussions the central importance that should be allocated to toxicokinetic considerations for the design and conduct of toxicological (*in vitro*) tests and the interpretation of the toxicity data generated by these tests.

Toxicokinetics is of particular importance for extrapolating from external to internal exposure and from *in vitro* data to the human *in vivo* situation. Full integration of toxicokinetic considerations into the planning and performing of toxicity testing was therefore identified as essential. It is imperative that all factors which could affect the derivation of biologically important concentrations from the *in vitro* experiments should be taken into consideration and be incorporated into the study design if judged or anticipated to be significant.¹⁶

The experts came up with a conceptual framework, as illustrated in Fig. 11, for the assessment of systemic toxicological effects without using animals as model for the human body.

In practical terms, the following is proposed:

First, information on the extent of exposure needs to be considered, taking into account exposure scenarios for multiple routes and multiple sources. Then the likely

exposure(s) need to be compared with the route-specific threshold of toxicological concern (TTC) value(s) (if these are known), i.e. exposure values below which no adverse effects have to be expected. If the external exposure is below the TTC value, the conclusion would be that the considered use of the substance is safe and a refined risk assessment, e.g. including animal testing, is not needed. If the external exposure is above the route-specific TTC value, then the internal or target dose, to which the target cells or tissues/organs might be exposed, has to be estimated.

For this estimation, knowledge on absorption, distribution, metabolism and excretion (ADME) of the substance under consideration becomes important. Most of this information can be deduced from appropriate *in vitro* and/or *in silico* studies, which will allow the estimation of the internal or target dose by means of physiologically based toxicokinetic (PBTK) modelling. PBTK models would be used as a means to convert external doses into internal concentration–time profiles and/or to convert concentrations used in the *in vitro* tests, which represent an assumed target dose, into external doses. If the internal dose (e.g. concentration of the target substance or its metabolites in the blood) is below the internal TTC value (determined from an appropriate database of *in vitro* assays), the conclusion would be that the assessed product use is safe and further testing is not needed.

It should be noted though that whereas the TTC approach for external exposure is becoming widely

¹⁶ Pelkonen et al. (2008a).

accepted particularly for food contaminants due to the available databases, the internal TTC approach, while valid as a concept, needs further data acquisition for providing a scientific solid basis.

If the internal TTC is not known, it is necessary to assess the potential impact of the estimated target dose on the target tissue/cells. This can be verified by means of exposing appropriate cell or tissue cultures (in vitro) to the target dose range or by carrying out in silico evaluations. At a minimum, the in vitro methods will allow a dose response relationship to be established, which would indicate if the estimated internal dose (range) is indeed of potential toxicological concern, i.e. able to induce effects at molecular, cellular or tissue level.

Using appropriate modelling techniques that take into consideration the relevant toxicokinetics, it may then be possible to extrapolate from the observed molecular, cellular or tissue level effects to the toxicological impact, i.e. the adverse health effect or disease at organism level. PBTK models will allow to link dose levels used in the in vitro experiments (and hence the related observed effects) with external exposure levels. However, it has to be pointed out that the relationship between observed effects in vitro to an adverse health effect at organism level remains to be established in many cases (area outside the shaded boxes in Fig. 11), and this could be the main rate-limiting step to the application of this framework in practice.

References

- Aardema MJ, Barnett B, Khambatta ZS, Reisinger K, Ouedrago-Arras G, Faquet B, Ginestet A, Mun GC, Dahl EL, Hewitt NJ, Corvi R, Curren RC (2010) International pre-validation studies of the EpiDerm™ 3D human reconstructed skin micronucleus (RSMN) assay: transferability and reproducibility. *Mutat Res* 701:123–131
- Adler S, Lindqvist J, Uddenberg K, Hyllner J, Strehl R (2008a) Testing potential developmental toxicants with a cytotoxicity assay based on human embryonic stem cells. *Altern Lab Anim* 36:129–140
- Adler S, Pellizzer C, Hareng L, Hartung T, Bremer S (2008b) First steps in establishing a developmental toxicity test method based on human embryonic stem cells. *Toxicol In Vitro* 22:200–211
- Adler S, Bicker G, Bigalke H, Bishop C, Blümel J, Dressler D, Fitzgerald J, Gessler F, Heuschen H, Kegel B, Luch A, Milne C, Pickett A, Ratsch H, Ruhdel I, Sesardic D, Stephens M, Stiens G, Thornton PD, Thürmer R, Vey M, Spielmann H, Grune B, Liebsch M (2010) The current scientific and legal status of alternative methods to the LD50 test for botulinum neurotoxin potency testing. The report and recommendations of a ZEBET expert meeting. *Altern Lab Anim* 38(4):315–330
- Aeby P, Python F, Goebel C (2007) Skin sensitisation: understanding the in vivo situation for the development of reliable in vitro test approaches. *ALTEX* 24(spec no):3–5
- Agdeppa ED, Spilker ME (2009) A review of imaging agent development. *AAPS J* 11(2):286–299
- Aleksic M, Thain E, Roger D, Saib O, Davies M, Li J, Aptula A, Zazzeroni R (2009) Reactivity profiling: covalent modification of single nucleophile peptides for skin sensitisation risk assessment. *Toxicol Sci* 108(2):401–411
- Aliahmadi E, Gramlich R, Grützkau A, Hitzler M, Krüger M, Baumgrass R, Schreiner M, Wittig B, Wanner R, Peiser M (2009) TLR2-activated human langerhans cells promote Th17 polarization via IL-1beta, TGF-beta and IL-23. *Eur J Immunol* 39(5):1221–1230
- Allen DD, Geldenhuys WJ (2006) Molecular Modelling of blood-barrier nutrient transporters: in silico basis for evaluation of potential drug delivery to the central nervous system. *Life Sci* 78:1029–1033
- Allen JW, Khetani SR, Bhatia SN (2005) In vitro zonation and toxicity in a hepatocyte bioreactor. *Toxicol Sci* 84(1):110–119
- Allen BC, Hack CE, Clewell HJ (2007) Use of Markov Chain Monte Carlo analysis with a physiologically-based pharmacokinetic model of methylmercury to estimate exposures in US women of childbearing age. *Risk Anal* 27:947–959
- Ames BN, Gold LS (1991) Response. *Science* 251:607–608
- Andersen ME (1995) What do we mean by dose? *Inhalation Toxicol* 7:909–915
- Andersen KE, Maibach HI (1985) Guinea pig sensitisation assays. An overview. In: Andersen KE, Maibach HI (eds) *Contact allergy. Predictive tests in guinea pigs*. Karger, Basel, pp 263–290
- Andersen KE, Vølund AA, Frankild S (1995) The guinea pig maximization test—with a multiple dose design. *Acta Derm Venereol* 75(6):463–469
- Aninat C, Piton A, Glaise D, Le Charpentier T, Languouët S, Morel F, Guguen-Guillouzo C, Guillouzo A (2006) Expression of cytochromes P450, conjugating enzymes and nuclear receptors in human hepatoma RG cells. *Drug Metab Dispos* 34(1):75–83
- Anon (1967) Report of the WHO scientific group. WHO principles for the testing of drugs for teratogenicity. WHO technical reports series 18, p 364
- Anon (1998) American society for testing and materials (ASTM). Standard guide for conducting the frog embryo teratogenesis assay—xenopus (FETAX), E1439-91 annual book of ASTM standards. ASTM, Philadelphia, 11.05, pp 826–836
- Anon (2003) Globally harmonised system of classification and labelling of chemicals (GHS). Part 3: health and environmental hazards. United Nations Organisation, New York, pp 151–158
- Anon (2010a) The official ReProTect Website. <http://www.reprotect.eu/>. Accessed on 12 July 2010
- Anon (2010b) Embryotoxicity testing in post-implantation embryo culture—method of piersma INVITTOX no 123. http://www.ecvam-dbalm.jrc.ec.europa.eu/public_view_doc2.cfm?id=15BF E9AC6ECC8ED3D7F6DF03B5D5BBE97180BB0BC12CB10496CDA74B54630A05A3291B895581F634. Accessed on 13 July 2010
- Ao L, Liu JY, Liu WB, Gao LH, Hu R, Fang ZJ, Zhen ZX, Huang MH, Yang MS, Cao J (2010) Comparison of gene expression profiles in BALB/c 3T3 transformed foci exposed to tumor promoting agents. *Toxicol In Vitro* 24:430–438
- Api AM, Basketter DA, Cadby PA, Cano MF, Ellis G, Gerberick GF, Griem P, McNamee PM, Ryan CA, Safford B (2008) Dermal sensitisation quantitative risk assessment (QRA) for fragrance ingredients. *Regul Toxicol Pharmacol* 52(1):3–23
- Aptula AO, Roberts DW (2006) Mechanistic applicability domains for non animal-based prediction of toxicological end points: general principles and application to reactive toxicity. *Chem Res Toxicol* 19(8):1097–1105
- Aptula AO, Roberts DW, Pease C (2007) Haptens, prohaptens and prehapten, or electrophiles and proelectrophiles. *Contact Derm* 56(1):54–56
- Artola-Garicano E, Vaes WH, Hermens JL (2000) Validation of negligible depletion solid-phase microextraction as a tool to

- determine tissue/blood partition coefficients for semivolatile and nonvolatile organic chemicals. *Toxicol Appl Pharmacol* 166:138–144
- Aschauer L, Wilmes A, Leonard MO, Pfaller W, Jennings P (2010) In vitro perfusion system for semi-automation of long-term repeat dose toxicity testing. *ALTEX* 27(suppl 2/10):12–13
- Ashby J (2001) Expectations for transgenic rodent cancer bioassay models. *Toxicol Pathol* 29(suppl):177–182
- Ashikaga T, Yoshida Y, Hirota M, Yoneyama K, Itagaki H, Sakaguchi H, Miyazawa M, Ito Y, Suzuki H, Toyoda H (2006) Development of an in vitro skin sensitisation test using human cell lines: the human cell line activation test (h-CLAT). I. Optimization of the h-CLAT protocol. *Toxicol In Vitro* 20(5):767–773
- Attardi BJ, Hild SA, Reel JR (2006) Dimethandrolone undecanoate: a new potent orally active androgen with progestational activity. *Endocrinology* 147:3016–3026
- Augustine KA (2009) HESI DART. Alternative assays workshop II: generation and preliminary assessment of a zebrafish teratogenicity assay. http://www.hesiglobal.org/files/public/Committees/DART/Meetings%20and%20Workshops/AlternativeAssayWkshp_Augustine.pdf. Accessed on 11 July 2010
- Augustine-Rauch K, Zhang CX, Panzica-Kelly JM (2010) In vitro developmental toxicology assays: a review of the state of the science of rodent and zebrafish whole embryo culture and embryonic stem cell assays. *Birth Defects Res C Embryo Today* 90:87–98
- Bailey JKABJ (2005) The future of teratology is in vitro. *Biogenic Amines* 19:97–145
- Baillie TA (2006) Future of toxicology-metabolic activation and drug design: challenges and opportunities in chemical toxicology. *Chem Res Toxicol* 19:889–893
- Baillie TA (2008) Metabolism and toxicity of drugs. Two decades of progress in industrial drug metabolism. *Chem Res Toxicol* 21:129–137
- Balls M, Hellsten E (2002) Statement of the scientific validity of the embryonic stem cell test (EST)—an in Vitro test for embryotoxicity. *Altern Lab Anim* 30:265–268
- Bal-Price A, Suñol C, Dieter G, Weiss G, van Vliet E, Westerink RHS, Costa LG (2008) Application of in vitro neurotoxicity testing for regulatory purposes: symposium III summary and research needs. *Neurotoxicology* 29(3):520–531
- Banas A, Teratani T, Yamamoto Y, Tokuhara M, Takeshita F, Quinn G, Okochi H, Ochiya T (2007) Adipose tissue-derived mesenchymal stem cells as a source of human hepatocytes. *Hepatology* 46(1):219–228
- Bantle JA, Finch RA, Fort DJ, Stover EL, Hull M, Kumsher-King M, Gaudet-Hull AM (1999) Phase III interlaboratory study of FETAX. Part 3. FETAX validation using 12 compounds with and without an exogenous metabolic activation system. *J Appl Toxicol* 19:447–472
- Barlow S (2005) Threshold of toxicological concern (TTC). A tool for assessing substances of unknown toxicity present at low levels in the diet. ILSI Europe concise monograph series, The International Life Sciences Institute (ILSI), Europe
- Barratt MD, Castell JV, Chamberlain M, Combes RD, Dearden JC, Fentem JH, Gerner I, Giuliani A, Gray TJB, Livingstone DJ, Provan WML, Rutten F, Verhaar HJM, Zbinden P (1995) The integrated use of alternative approaches for predicting toxic hazard—the report and recommendations of ECVAM workshop-8. *ATLA* 23:410–429
- Barton HA (2009) Biological modeling as a method for data evaluation and integration in toxicology. *Hum Exp Toxicol* 28(2–3):143–145
- Barton HA, Pastoor TP, Baetcke K, Chambers JE, Diliberto J, Doerrer NG, Driver JH, Hastings CE, Iyengar S, Krieger R, Stahl B, Timchalk C (2006) The acquisition and application of absorption, distribution, metabolism, and excretion (ADME) data in agricultural chemical safety assessments. *Crit Rev Toxicol* 36(1):9–35
- Barton HA, Chiu WA, Setzer RW, Andersen ME, Bailer AJ, Bois FY, Dewoskin RS, Hays S, Johanson G, Jones N, Loizou G, Macphail RC, Portier CJ, Spendiff M, Tan YM (2007) Characterizing uncertainty and variability in physiologically based pharmacokinetic models: state of the science and needs for research and implementation. *Toxicological*
- Basketter DA (2009) The human repeated insult patch test in the 21st century: a commentary on ethics and validity. *Cutan Ocul Toxicol* 28(2):49–53
- Basketter DA (2010) Methyl dibromoglutaronitrile, skin sensitisation and quantitative risk assessment. *Cutan Ocul Toxicol* 29(1):4–9
- Basketter DA, Kimber I (2010a) Contact hypersensitivity. In: McQueen CA (ed) *Comprehensive toxicology*, vol 5, 2nd edn edn. Elsevier, Kidlington, pp 397–411
- Basketter DA, Kimber I (2010b) Skin sensitization, false positives and false negatives: experience with guinea pig assays. *J Appl Toxicol* 30(5):381–386
- Basketter DA, Lea LJ, Dickens A, Briggs D, Pate I, Dearman RJ, Kimber I (1999) A comparison of statistical approaches to the derivation of EC3 values from local lymph node assay dose responses. *J Appl Toxicol* 19(4):261–266
- Basketter DA, Andersen KE, Lidén C, van Loveren H, Boman A, Kimber I, Alanko K, Berggren E (2005a) Evaluation of the skin sensitising potency of chemicals using existing methods and considerations of relevance for elicitation. *Contact Derm* 52(1):39–43
- Basketter DA, Casati S, Gerberick GF, Griem P, Philips B, Worth A (2005b) Skin Sensitisation. *Altern Lab Anim* 33(suppl 1):83–103
- Basketter DA, Clapp C, Jefferies D, Safford B, Ryan CA, Gerberick GF, Dearman RJ, Kimber I (2005c) Predictive identification of human skin sensitisation thresholds. *Contact Derm* 53(5):260–267
- Basketter DA, Gerberick GF, Kimber I (2007a) The local lymph node assay EC3 value: status of validation. *Contact Derm* 57(2):70–75
- Basketter DA, Kan-King-Yu D, Dierkes P, Jowsey IR (2007b) Does irritation contribute to the skin sensitisation potency of contact allergens? *Cutan Ocul Toxicol* 26(4):279–286
- Basketter DA, Casati S, Cronin MTD, Diembeck W, Gerberick GF, Hadgraft J, Kasting G, Marty JP, Nikolaidis E, Patlewicz G, Pease C, Roberts DW, Roggen E, Rovida C, van der Sandt J (2007c) Skin sensitisation and epidermal disposition. *Altern Lab Anim* 35(1):137–154
- Basketter DA, Darlenski R, Fluhr J (2008a) Skin irritation and sensitisation: mechanisms and new approaches for risk assessment. *Skin Pharmacol Physiol* 21(4):191–202
- Basketter DA, Clapp CJ, Safford BJ, Jowsey IR, McNamee PM, Ryan CA, Gerberick GF (2008b) Preservatives and skin sensitisation quantitative risk assessment: risk benefit considerations. *Dermatitis* 19(1):20–27
- Basketter DA, Kimber I (2009) Updating the skin sensitisation in vitro data assessment paradigm in 2009. *J Appl Toxicol* 29(6):545–550
- Battelle (2002) Revised AR binding assay protocol: androgen receptor competitive binding protocol using rat ventral prostate cytosol, SOP no. NHEERL-H/RTD/EB/VW/2002-03-000. In: Quality assurance project plan for work assignment 2-22, Development of Androgen Receptor Binding Data, prepared by Battelle for US environmental protection agency, endocrine disruptor screening program, contract no. 68-W-01-023
- Baylin SB, Ohm JE (2006) Epigenetic gene silencing in cancer—a mechanism for early oncogenic pathway addition? *Nat Rev Cancer* 6:107–116

- Benfenati E, Benigni R, DeMarini DM, Helma C, Kirkland D, Martin TM, Mazzatorta P, Ouedraogo-Arras G, Richard AM, Schilter B, Schoonen WGEJ, Snyder RD, Yang C (2009) Predictive models for carcinogenicity and mutagenicity: frameworks, state-of-the-art, and perspectives. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev* 27(2):57–90
- Benigni R, Bossa C (2008) Predictivity of QSAR. *J Chem Inf Model* 48:971–980
- Benigni R, Bossa C, Tcheremenskaia O, Giuliani A (2010) Alternatives to the carcinogenicity bioassay: in silico methods, and the in vitro and in vivo mutagenicity assays. *Expert Opin Drug Metab Toxicol* 6:809–819
- Bercu JP, Morton SM, Deahl JT, Gombar VK, Callis CM, van Lier RB (2010) In silico approaches to predicting cancer potency for risk assessment of genotoxic impurities in drug substances. *Regul Toxicol and Pharmacol* 57:300–306
- Bernillon P, Bois FY (2000) Statistical issues in toxicokinetic modeling: a Bayesian perspective. *Environ Health Perspect* 108:883–893
- Bernshausen T, Noss A, Boehn SN, Fabian E, Kamp HG, Kaufmann T, Koch M, Wessa P, van Ravenzwaay B, Landsiedel R (2009) Metabolic Competences of the Chicken Embryo used in the Toxicity Screening Test (CHEST) as an In Vitro Test System for Teratogenicity and Embryotoxicity. 50. Jahrestagung der DGPT, Mainz, Germany
- Bérubé K, Aufderheide M, Breheny D, Clothier R, Combes R, Duffin R, Forbes B, Gaça M, Gray A, Hall I, Kelly M, Lethem M, Liebsch M, Merolla L, Morin JP, Seagrave JC, Swartz MA, Tetley TD, Umachandran M (2009) In Vitro Models of Inhalation Toxicity and Disease. The Report of a FRAME Workshop. *Altern Lab Anim* 37(1):89–141
- Bessemers JGM (2009) Opinion on the usefulness of in vitro data for human risk assessment. Suggestions for better use of non-testing approaches. Letter report 320016002/2009 RIVM. <http://www.rivm.nl/bibliotheek/rapporten/320016002.pdf>. Accessed 15 January 2011
- Billinton N, Hastwell PW, Beerens D, Birrell L, Ellis P, Maskell S, Webster TW, Windebank S, Woestenborghs F, Lynch AM, Scott AD, Tweats DJ, van Gompel J, Rees RW, Walmsley RM (2008) Interlaboratory assessment of the GreenScreen HC GADD45a-GFP genotoxicity screening assay: an enabling study for independent validation as an alternative method. *Mutat Res* 653:23–33
- Billinton N, Bruce S, Hansen JR, Hastwell PW, Jagger C, McComb C, Klug ML, Pant K, Rabinowitz A, Rees R, Tate M, Vinggaard AM, Walmsley RM (2010) A pre-validation transferability study of the GreenScreen HC GADD45a-GFP assay with a metabolic activation system (S9). *Mutat Res* 700:44–50
- Bitsch A, Jacobi S, Melber C, Wahnschaffe U, Simetska N, Mangelsdorf I (2006) REPDOSE: a database on repeated dose toxicity studies of commercial chemicals—a multifunctional tool. *Regul Toxicol Pharmacol* 46(3):202–210
- Bittner M, Hilscherova K, Giesy JP (2009) In vitro assessment of AhR-mediated activities of TCDD in mixture with humic substances. *Chemosphere* 76:1505–1508
- Blaauboer BJ (2002) The necessity of biokinetic information in the interpretation of in vitro toxicity data. *Altern Lab Anim* 30(suppl 2):85–91
- Blaauboer BJ (2003) The integration of data on physico-chemical properties, in vitro-derived toxicity data and physiologically based kinetic and dynamic modelling as a tool in hazard and risk assessment. A commentary. *Toxicol Lett* 138(1–2):161–171
- Blaauboer BJ (2008) The contribution of in vitro toxicity data in hazard and risk assessment: current limitations and future perspectives. *Toxicol Lett* 180(2):81–84
- Blaauboer BJ (2010) Biokinetic modeling and in vitro-in vivo extrapolations. *J Toxicol Environ Health B Crit Rev* 13(2–4):242–252
- Blaauboer BJ, Andersen ME (2007) The need for a new toxicity testing and risk analysis paradigm to implement REACH or any other large scale testing initiative. *Arch Toxicol* 81(5):385–387
- Blaauboer BJ, Bayliss MK, Castell JV, Evelo CTA, Frazier JM, Groen K, Gulden M, Guillouzo A, Hissink AM, Houston BJ, Johanson G, de Jongh J, Kedderis GL, Reinhardt CA, van de Sandt JJM, Semino G (1996) The use of biokinetics and in vitro methods in toxicological risk evaluation. *ATLA* 24:473–497
- Blaauboer BJ, Barratt MD, Houston BJ (1999) The integrated use of alternative methods in toxicological risk evaluation: ECVAM integrated testing strategies task force report 1. *ATLA* 27:229–237
- Blackburn K, Stickney JA, Carlson-Lynch HL, McGinnis PA, Chappell L, Felter SP (2005) Application of the threshold of toxicological concern approach to ingredients in personal and household care products. *Regul Toxicol and Pharmacol* 43:249–259
- Boehn SNE, Kaufmann T, Koch M, Wessa P, Kamp HG, Landsiedel R, van Ravenzwaay B (2009) Modification and evaluation of the chicken embryotoxicity screening test (CHEST) as in vitro test system for embryotoxicity. VII World Congress on Alternatives & Animal Use in the Life Sciences, Rome
- Boekelheide K, Campion SN (2010) Toxicity testing in the 21st century: using the new toxicity testing paradigm to create a taxonomy of adverse effects. *Toxicol Sci* 114(1):20–24
- Boess F, Kamber M, Romer S, Gasser R, Muller D, Albertini S, Suter L (2003) Gene expression in two hepatic cell lines, cultured primary hepatocytes, and liver slices compared to the in vivo liver gene expression in rats: possible implications for toxicogenomics use of in vitro systems. *Toxicol Sci* 73(2):386–402
- Böhme A, Thaens D, Paschke A, Schüürmann G (2009) Kinetic glutathione chemoassay to quantify thiol reactivity of organic electrophiles application to α,β -unsaturated ketones, acrylates, and propiolates. *Chem Res Toxicol* 22(4):742–750
- Bois FY, Paxman D (1992) An analysis of exposure rate effects for benzene using a physiologically based pharmacokinetic model. *Regul Toxicol Pharmacol* 15:122–136
- Bois FY, Smith T, Gelman A, Chang HY, Smith AE (1999) Optimal design for a study of butadiene toxicokinetics in humans. *Toxicol Sci* 49:213–224
- Bois FY, Jamei M, Clewell HJ (2010) PBPK modelling of inter-individual variability in the pharmacokinetics of environmental chemicals. *Toxicology* 278(3):256–267
- Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, Vickers C, Willcocks D, Farland W (2006) IPCS framework for analyzing the relevance of a cancer mode of action for humans. *Crit Rev Toxicol* 36(10):781–792
- Boobis AR, Doe JE, Heinrich-Hirsch B, Meek ME, Munn S, Ruchirawat M, Schlatter J, Seed J, Vickers C (2008) IPCS framework for analyzing the relevance of a noncancer mode of action for humans. *Crit Rev Toxicol* 38(2):87–96
- Boobis AR, Daston GP, Preston RJ, Olin SS (2009a) Application of key events analysis to chemical carcinogens and noncarcinogens. *Crit Rev Food Sci Nutr* 49(8):690–707
- Boobis AR, Cohen SM, Doerr NG, Galloway SM, Haley PJ, Hard GC, Hess FG, Macdonald JS, Thibault S, Wolf DC, Wright J (2009b) A data-based assessment of alternative strategies for identification of potential human cancer hazards. *Toxicol Pathol* 37(6):714–732
- Boogaard PJ, Nagelkerke JF, Mulder GJ (1990) Renal proximal tubular cells in suspension or in primary culture as in vitro models to study nephrotoxicity. *Chem Biol Interact* 76(3):251–291
- Bouvier d'Yvoire M, Prieto P, Blaauboer BJ, Bois FY, Boobis A, Brochot C, Coecke S, Freidig A, Gundert-Remy U, Hartung T, Jacobs MN, Lav T, Leahy DE, Lennernäs H, Loizou GD, Meek

- B, Pease C, Rowland M, Spendiff M, Yang J, Zeilmaker M (2007) Physiologically-based kinetic modelling (PBK modelling): meeting the 3Rs agenda: the report and recommendations of ECVAM workshop 63. *ATLA* 35:661–671
- Bouwman T, Cronin MT, Bessems JG, van de Sandt JJ (2008) Improving the applicability of (Q)SARs for percutaneous penetration in regulatory risk assessment. *Hum Exp Toxicol* 27:269–276
- Brandenberger C, Rothen-Rutishauser B, Mühlfeld C, Schmid O, Ferron GA, Maier KL, Gehr P, Lenz AG (2010) Effects and uptake of gold nanoparticles deposited at the air–liquid interface of a human epithelial airway model. *Toxicol Appl Pharmacol* 242(1):56–65
- Brandon EF, Oomen AG, Rempelberg CJ, Versantvoort CH, van Engelen JG, Sips AJ (2006) Consumer product in vitro digestion model: bioaccessibility of contaminants and its application in risk assessment. *Regul Toxicol Pharmacol* 44:161–171
- Brannen KC, Panzica-Kelly JM, Danberry TL, Augustine-Rauch KA (2010) Development of a zebrafish embryo teratogenicity assay and quantitative prediction model. *Birth Defects Res B Dev Reprod Toxicol* 89:66–77
- Bremer S, Van Dooren M, Paparella M, Kossolov E, Fleischmann B, Hescheler J (1999) Establishment of an embryotoxicity assay with green fluorescence protein-expressing embryonic cell-derived cardiomyocytes. *Atla Altern Lab Anim* 27:471–484
- Bremer S, Pellizzer C, Hoffmann S, Seidle T, Hartung T (2007) The development of new concepts for assessing reproductive toxicity applicable to large scale toxicological programmes. *Curr Pharm Des* 13:3047–3058
- Brown NA, Fabro S (1981) Quantitation of rat embryonic development in vitro: a morphological scoring system. *Teratology* 24:65–78
- Brown NA, Spielmann H, Bechter R, Flint OP, Freeman SJ, Jelinek RJ, Koch E, Nau H, Newall DR, Palmer AK et al (1995) Screening chemicals for reproductive toxicity: the current alternatives; the report and recommendations of an ECVAM/ETS workshop (ECVAM workshop 12). *Altern Lab Anim* 23(6):868–882
- Broxmeyer HE, Srour E, Orschell C, Ingram DA, Cooper S, Plett PA, Mead LE, Yoder MC (2006) Cord blood stem and progenitor cells. *Methods Enzymol* 419:439–473
- Bucher JR (1998) Update on national toxicology program (NTP) assays with genetically altered or “transgenic” mice. *Environ Health Perspect* 106:619–621
- Buehler EV (1965) Delayed contact hypersensitivity in the guinea pig. *Arch Dermatol* 91:171–177
- Buesen R, Genschow E, Slawik B, Visan A, Spielmann H, Luch A, Seiler A (2009) Embryonic stem cell test remastered: comparison between the validated EST and the new molecular FACS-EST for assessing developmental toxicity in vitro. *Toxicol Sci* 108:389–400
- Busquet F, Nagel R, von Landenberg F, Mueller SO, Huebler N, Broschard TH (2008) Development of a new screening assay to identify proteratogenic substances using zebrafish *Danio rerio* embryo combined with an exogenous mammalian metabolic activation system (mDarT). *Toxicol Sci* 104:177–188
- Buzanska L, Sypecka J, Nerini-Molteni S, Compagnoni A, Hogberg HT, del Torchio R, Domanska-Janik K, Zimmer J, Coecke S (2009) A human stem cell-based model for identifying adverse effects of organic and inorganic chemicals on the developing nervous system. *Stem Cells* 27(10):2591–2601
- Caesar software (2010) <http://www.caesar-project.eu>. Accessed 30 June 2010
- Carfí M, Gennari A, Malerba I, Corsini E, Pallardy M, Pieters R, Van Loveren H, Vohr HW, Hartung T, Gribaldo L (2007) In vitro tests to evaluate immunotoxicity: a preliminary study. *Toxicology* 229(1–2):11–22
- Carfí M, Bowe G, Ferrario D, Pieters R, Gribaldo L (2010) Maintenance and characterization of lymphocytes in human long term bone marrow cultures to study immunotoxicity. *Toxicol In Vitro* 24(5):1395–1403
- Carmichael P, Davies M, Dent M, Fentem J, Fletcher S, Gilmour N, MacKay C, Maxwell G, Merolla L, Pease C, Reynolds F, Westmoreland C (2009) Non-animal approaches for consumer safety risk assessments: unilever’s scientific research programme. *Altern Lab Anim* 37(6):595–610
- Carthew P, Clapp C, Gutsell S (2009) Exposure based waiving: the application of the toxicological threshold of concern (TTC) to inhalation exposure for aerosol ingredients in consumer products. *Food Chem Toxicol* 47:1287–1295
- Casati S, Aeby P, Basketter DA, Cavani A, Gennari A, Gerberick GF, Griem P, Hartung T, Kimber I, Lepoittevin JP, Meade BJ, Pallardy M, Rougier N, Rousset F, Rubinstenn G, Sallusto F, Verheyen GR, Zuang V (2005) Report and recommendations of ECVAM workshop 51. Dendritic cells as a tool for the predictive identification of skin sensitisation hazard. *Altern Lab Anim* 33(1):47–62
- Cecchelli R, Berezowski V, Lundquist S, Culot M, Renftel M, Dehouck MP, Fenart L (2007) Modelling of the blood–brain barrier in drug discovery and development. *Nat Rev Drug Disc* 6(8):650–661
- Cefic (2010) <http://www.cefic-lri.org/lri-toolbox/ambit>. Accessed 30 June 2010
- Cezar GG, Quam JA, Smith AM, Rosa GJ, Piekarczyk MS, Brown JF, Gage FH, Muotri AR (2007) Identification of small molecules from human embryonic stem cells using metabolomics. *Stem Cells Dev* 16:869–882
- Chao P, Maguire T, Novik E, Cheng KC, Yarmush ML (2009) Evaluation of a microfluidic based cell culture platform with primary human hepatocytes for the prediction of hepatic. *Biochem Pharmacol* 78:625–632
- Chapin RE, Stedman DB (2009) Endless possibilities: stem cells and the vision for toxicology testing in the 21st century. *Toxicol Sci* 112(1):17–22
- Chapin R, Stedman D, Paquette J, Streck R, Kumpf S, Deng S (2007) Struggles for equivalence: in vitro developmental toxicity model evolution in pharmaceuticals in 2006. *Toxicol In Vitro* 21:1545–1551
- Chaudhry Q, Piclin N, Cotterill J, Pintore M, Price NR, Chrétien JR, Roncaglioni A (2010) Global QSAR models of skin sensitizers for regulatory purposes. *Chem Cent J* 4(suppl 1):S5
- Cheeseman MA, Machuga EJ, Bailey AB (1999) A tiered approach to threshold of regulation. *Food Chem Toxicol* 37:387–412
- Chipinda I, Ajibola RO, Morakinyo MK, Ruwona TB, Simoyi RH, Siegel PD (2010) Rapid and simple kinetics screening assay for electrophilic dermal sensitizers using nitrobenzenethiol. *Chem Res Toxicol* 23(5):918–925
- Choi S, Park SY, Jeong J, Cho E, Phark S, Lee M, Kwak D, Lim JY, Jung WW, Sul D (2010) Identification of toxicological biomarkers of di(2-ethylhexyl) phthalate in proteins secreted by HepG2 cells using proteomic analysis. *Proteomics* 10(9):1831–1846
- Clewell HJ, Tan YM, Campbell JL, Andersen ME (2008) Quantitative interpretation of human biomonitoring data. *Toxicol Appl Pharmacol* 231:122–133
- Coecke S, Rogiers V, Bayliss M, Castell J, Doehmer J, Fabre G, Fry J, Kern A, Westmoreland C (1999) The use of long-term hepatocyte cultures for detection induction of drug metabolising enzymes: the current status. *ATLA* 27:579–638
- Coecke S, Blaauboer BJ, Elaut G, Freeman S, Freidig A, Gensmantel N, Hoet P, Kapoulas VM, Ladstetter B, Langley G, Leahy D, Mannens G, Meneguz A, Monshouwer M, Nemery B, Pelkonen

- O, Pfaller W, Prieto P, Proctor N, Rogiers V, Rostami-Hodjegan A, Sabbioni E, Steiling W, van de Sandt JJ (2005) Toxicokinetics and metabolism. *Altern Lab Anim* 33(1):147–175
- Coecke S, Ahr H, Blaauboer BJ, Bremer S, Casati S, Castell J, Combes R, Corvi R, Crespi CL, Cunningham ML, Elaut G, Eletti B, Freidig A, Gennari A, Ghersi-Egea JF, Guillouzo A, Hartung T, Hoet P, Ingelman-Sundberg M, Munn S, Janssens W, Ladstetter B, Leahy D, Long A, Meneguz A, Monshouwer M, Morath S, Nagelkerke F, Pelkonen O, Ponti J, Prieto P, Richert L, Sabbioni E, Schaack B, Steiling W, Testai E, Vericat JA, Worth A (2006) Metabolism: a bottleneck in in vitro toxicological test development. The report and recommendations of ECVAM workshop 54. *ATLA* 34(1):49–84
- Cohen SM (2004) Human carcinogenic risk evaluation: an alternative approach to the 2-year rodent bioassay. *Toxicol Sci* 80(2):225–229
- Combes R, Balls M, Illing P, Bhogal N, Dale J, Duv' G, Feron V, Grindon C, Gulden M, Loizou G, Priston R, Westmoreland C (2006) Possibilities for a new approach to chemicals risk assessment. The report of a FRAME workshop. *Altern Lab Anim* 34(6):621–649
- Conolly RB, Butterworth BE (1995) Biologically based dose response model for hepatic toxicity: a mechanistically based replacement for traditional estimates of noncancer risk. *Toxicol Lett* 82–83: 901–906
- Constant S (2010) Assessment of acute, long-term and chronic respiratory toxicity using a long shelf-life 3D model of the human airway epithelium. *ALTEX* 27(suppl 2/10):29
- Contrera JF, Matthews EJ, Kruhlak NL, Benz RD (2004) Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modelling of the human maximum recommended daily dose. *Regul Toxicol Pharmacol* 40(3): 185–206
- Contrera JF, Kruhlak NL, Matthews EJ, Benz RD (2007) Comparison of MC4PC and MDL-QSAR rodent carcinogenicity predictions and the enhancement of predictive performance by combining QSAR models. *Regul Toxicol Pharmacol* 49:172–182
- Cooper RL, Lamb JC, Barlow SM, Bentley K, Brady AM, Doerrer NG, Eisenbrandt DL, Fenner-Crisp PA, Hines RN, Irvine LF, Kimmel CA, Koeter H, Li AA, Makris SL, Sheets LP, Speijers G, Whitby KE (2006) A tiered approach to life stages testing for agricultural chemical safety assessment. *Crit Rev Toxicol* 36:69–98
- Cordelli E, Freseghna AM, D'Alessio A, Eleuteri P, Spano M, Pacchierotti F, Villani P (2007) ReProComet: a new in vitro method to assess DNA damage in mammalian sperm. *Toxicol Sci* 99:545–552
- Corsini E, Mitjans M, Galbiati V, Lucchi L, Galli CL, Marinovich M (2009) Use of IL-18 production in a human keratinocyte cell line to discriminate contact sensitizers from irritants and low molecular weight respiratory allergens. *Toxicol In Vitro* 23(5):789–796
- Corvi R, Albertini S, Hartung T, Hoffmann S, Maurici D, Pfuhrer S, van Benthem J, Vanparys P (2008) ECVAM retrospective validation of in vitro micronucleus test (MNT). *Mutagenesis* 23:271–283
- Cramer GM, Ford RA, Hall RL (1978) Estimation of toxic hazard—a decision tree approach. *Food Cosmet Toxicol* 16:255–276
- Crean D, Aschauer L, Jennings P, Leonard M (2010) Toxicity profiling of nephrotoxins in the human proximal tubular RPTEC-TERT1 cell line: influence of hypoxia. *ALTEX* 27(suppl 2/10):31
- Creton S, Billington R, Davies W, Dent MP, Hawksworth GM, Parry S, Travis KZ (2009) Application of toxicokinetics to improve chemical risk assessment: implications for the use of animals. *Regul Toxicol Pharmacol* 55(3):291–299
- Crivori P, Poggesi I (2006) Computational approaches for predicting CYP-related metabolism properties in the screening of new drugs. *Eur J Med Chem* 41(7):795–808
- Cronin MTD, Hewitt M (2007) In silico models to predict passage through the skin and other barriers. In: Testa B, van de Waterbeemd H (eds) *Comprehensive medicinal chemistry II*. Elsevier, Oxford, pp 725–744
- Cronin MTD, Worth AP (2008) (Q)SARs for predicting effects relating to reproductive toxicity. *Qsar Combinat Sci* 27:91–100
- Curren RD, Southee JA, Spielmann H, Liebsch M, Fentem JH, Balls M (1995) The role of pre-validation in the development, validation and acceptance of alternative methods. *Altern Lab Anim* 23:211–217
- Curren RD, Mun GC, Gibson DP, Aardema MJ (2006) Development of a method for assessing micronucleus induction in a 3D human skin model EpiDermTM. *Mutat Res* 607:192–204
- Dahl SG, Aarons L, Gundert-Remy U, Karlsson MO, Schneider YJ, Steimer JL, Trocéniz IF (2010) Incorporating physiological and biochemical mechanisms into pharmacokinetic-pharmacodynamic models: a conceptual framework. *Basic Clin Pharmacol Toxicol* 106:2–12
- Davies M, Pendlington RU, Page L, Roper CS, Sanders DJ, Bourner C, Pease CK, Mackay C (2010) Determining epidermal disposition kinetics for use in an integrated non-animal approach to skin sensitisation risk assessment. *Toxicol Sci* kfq326v1–kfq326
- De Bruyn AMH, Gobas FAPC (2007) The sorptive capacity of animal protein. *Environ Toxicol Chem* 26:1803–1808
- De Buck SS, Sinha VK, Fenu LA, Gilissen RA, Mackie CE, Nijssen MJ (2007) The prediction of drug metabolism, tissue distribution, and bioavailability of 50 structurally diverse compounds in rat using mechanism-based absorption, distribution, and metabolism prediction tools. *Drug Metab Dispos* 35:649–659
- De Gottardi A, Vinciguerra M, Sgroi A, Moukil M, Ravier-Dall'Antonia F, Paziienza V, Pugnale P, Foti M, Hadengue A (2007) Microarray analyses and molecular profiling of steatosis induction in immortalized human hepatocytes. *Lab Invest* 87(8):792–806
- de Graaf C, Vermeulen NP, Feenstra KA (2005) Cytochrome P450 in silico: an integrative modeling approach. *J Med Chem* 48(8):2725–2755
- De Kock J, Vanhaecke T, Biernaskie JA, Rogiers V, Snykers S (2009) Characterization and hepatic differentiation of skin-derived precursors from adult foreskin by sequential exposure to hepatogenic cytokines and growth factors reflecting liver development. *Toxicol In Vitro* 23(8):1522–1527
- de Longueville F, Atienzar FA, Marcq L, Dufrane S, Evrard S, Wouters L, Leroux F, Bertholet V, Gerin B, Whomsley R, Arnould T, Remacle J, Canning M (2003) Use of a low-density microarray for studying gene expression patterns induced by hepatotoxicants on primary cultures of rat hepatocytes. *Toxicol Sci* 75(2):378–392
- De Vries A, van Steeg H, Opperhuizen A (2004) Transgenic mice as alternatives in carcinogenicity testing: current status. *RIVM report* 340700001
- Dearden JC, Barratt MD, Benigni R, Bristol DW, Combes RD, Cronin MTD, Judson PN, Payne MP, Richard AM, Tichy M, Worth AP, Yourick JJ (1997) The development and validation of expert systems for predicting toxicity. *Altern Lab Anim* 25: 223–252
- DeJongh J, Forsby A, Houston JB, Beckman M, Combes R, Blaauboer BJ (1999a) An integrated approach to the prediction of systemic toxicity using computer-based biokinetic models and biological in vitro test methods: overview of a prevalidation study based on the ECITTS project. *Toxicol In Vitro* 13:549–554
- DeJongh J, Nordin-Andersson M, Ploeger BA, Forsby A (1999b) Estimation of systemic toxicity of acrylamide by integration on

- in vitro toxicity data with kinetic simulations. *Toxicol Appl Pharmacol* 158(3):261–268
- Deschl U, Vogel J, Aufderheide M (2010) Development of an in vitro exposure model for investigating the biological effects of therapeutic aerosols on human cells from the respiratory tract. *Exp Toxicol Pathol*. doi:10.1016/j.etp.2010.04.013
- Devillers J (2009) Endocrine disruption modeling. CRC Press, Boca Raton
- DiPaolo JA, Donovan P, Nelson R (1969) Quantitative studies of in vitro transformation by chemical carcinogens. *J Natl Cancer Inst* 42:867–874
- Divkovic M, Pease CM, Gerberick GF, Basketter DA (2005) Hapten-protein binding: from theory to practical application in the in vitro prediction of skin sensitisation. *Contact Derm* 53(4):189–200
- Donato MT, Martínez-Romero A, Jiménez N, Negro A, Herrera G, Castell JV, O'Connor JE, Gómez-Lechón MJ (2009) Cytometric analysis for drug-induced steatosis in HepG2 cells. *Chem Biol Interact* 181(3):417–423
- Doull J, Cattley R, Elcombe C, Lake BG, Swenberg J, Wilkinson C, Williams G, van Gemert M (1999) A cancer risk assessment of di(2-ethylhexyl) phthalate: application of the new US EPA risk assessment guidelines. *Regul Toxicol Pharmacol* 29:327–357
- Dr. Hadwen Trust Science Review (2006) Showcasing successes in non-animal research. In Dr. Hadwen Trust Science Review. Dr. Hadwen Trust. <http://www.drhadwentrust.org/>
- Durmus E, Inan O, Celik I, Sur E, Ozkan Y, Acar A, Aydin MF (2005) Use of the fertilized hen's egg in the evaluation of embryotoxicity of dental alloys. *J Biomed Mater Res B Appl Biomater* 72:322–327
- Eastin WC, Haseman JK, Mahler JF, Bucher JR (1998) The national toxicology program evaluation of genetically altered mice as predictive models for identifying carcinogens. *Toxicol Pathol* 26:461–473
- ECHA (2008) Guidance on information requirements and chemical safety assessment. Chapter R.7a: endpoint specific guidance http://www.echa.europa.eu/about/reach_en.asp. Accessed 15 Jan 2011
- ECHA (2010) Practical guide on how to report categories and read-across. http://www.echa.europa.eu/doc/publications/practical_guides/pg_report_readacross_cate.pdf
- Edginton AN, Theil FP, Schmitt W, Willmann S (2008) Whole body physiologically-based pharmacokinetic models: their use in clinical drug development. *Expert Opin Drug Metab Toxicol* 4:1143–1152
- EFSA, European Food Safety Authority (2004) Scientific panel on food additives, flavourings, processing aids and materials in contact with food. Opinion on flavouring group FGE.03 Acetals of branched- and straight-chain aliphatic saturated primary alcohols and branched- and straight-chain saturated aldehydes, and an orthoester of formic acid, from chemical groups 1 and 2. Opinion expressed on 7 Oct 2004. http://www.efsa.eu.int/science/afc/catindex_en.html
- EFSA, European Food Safety Authority (2009) Report of the PPR Unit on the outcome of the public consultation on the revision of the existing guidance document on dermal absorption scientific opinion on the revision of the guidance document on dermal absorption
- El-Ali J, Sorger PK, Jensen KF (2006) Cells on chips. *Nature* 442:403–411
- Elaut G, Henkens T, Papeleu P, Snykers S, Vinken M, Vanhaecke T, Rogiers V (2006) Molecular mechanisms underlying the dedifferentiation process of isolated hepatocytes and their cultures. *Curr Drug Metab* 7(6):629–660
- Elferink MG, Olinga P, Draaisma AL, Merema MT, Bauerschmidt S, Polman J, Schoonen WG, Groothuis GM (2008) Microarray analysis in rat liver slices correctly predicts in vivo hepatotoxicity. *Toxicol Appl Pharmacol* 229(3):300–309
- Ellinger-Ziegelbauer H, Gmuender H, Bandenburg A, Ahr HJ (2008) Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. *Mutat Res* 637:23–39
- Ellinger-Ziegelbauer H, Aubrecht J, Kleinjans JC, Ahr HJ (2009a) Application of toxicogenomics to study mechanisms of genotoxicity and carcinogenicity. *Toxicol Lett* 186:36–44
- Ellinger-Ziegelbauer H, Fostel JM, Aruga C, Bauer D, Boitier E, Deng S, Dickinson D, Le Fevre AC, Fornace AJ Jr, Grenet O, Gu Y, Hoflack JC, Shiiyama M, Smith R, Snyder RD, Spire C, Tanaka G, Aubrecht J (2009b) Characterization and interlaboratory comparison of a gene expression signature for differentiating genotoxic mechanisms. *Toxicol Sci* 10:341–352
- EMA (2006) European medicines evaluation agency. Committee for medicinal products for human use (CHMP). Guideline on the limits of genotoxic impurities. CPMP/SWP/5199/02. London, 28 June 2006. EMA/CHMP/QWP/251344/2006. <http://www.emea.europa.eu/pdfs/human/swp/519902en.pdf>
- EMA (2008) Committee for medicinal products for human use (CHMP). Questions & answers on the CHMP guideline on the limits of genotoxic impurities. Rev 1. London, 26 June 2008. Doc.Ref. EMA/CHMP/SWP/431994/2007
- Emter R, Ellis G, Natsch A (2010) Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. *Toxicol Appl Pharmacol* 245(3):281–290
- Enoch SJ, Cronin MTD (2010) A review of the electrophilic reaction chemistry involved in covalent DNA binding. *Crit Rev Toxicol* 40:728–748
- Enoch SJ, Madden JC, Cronin MT (2008) Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR QSAR Environ Res* 19(5–6):555–578
- Enoch SJ, Cronin MTD, Madden JC, Hewitt M (2009) Formation of structural categories to allow for read-across for teratogenicity. *QSAR Combinat Sci* 28:696–708
- Enslin K, Gombar VK, Blake BW (1994) International commission for protection against environmental mutagens and carcinogens. Use of SAR in computer-assisted prediction of carcinogenicity and mutagenicity of chemicals by the TOPKAT program. *Mutat Res* 205:47–61
- Eriksson L, Jaworska JS, Worth AP, Cronin MTD, McDowell RM, Gramatica P (2003) Methods for reliability, uncertainty assessment, and applicability evaluations of classification and regression based QSARs. *Environmental Health Perspectives* 22:1361–1375 <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1241620/?tool=pubmed>
- Escher SE, Tluczkiewicz I, Batke M, Bitsch A, Melber C, Kroese ED, Buist HE, Mangelsdorf I (2010) Evaluation of inhalation TTC values with the database RepDose. *Regul Toxicol Pharmacol* 58:259–274
- Eskes C, Zuang V (2005) Alternative (non-animal) methods for cosmetics testing: current status and future prospects. A report prepared in the context of the 7th amendment of the cosmetics directive for establishing the timetable for phasing out animal testing. *ATLA* 33(suppl 1):19–20
- Esteller M (2007) Cancer epigenomics. DNA methylomes and histone-modification maps. *Nat Rev Genet* 8:286–298
- EU (1976) Council directive 76/768/EEC of July 1976 on the approximation of the laws of the member states relating to cosmetic products. *Off J L262*:169–200
- EU (2003) Directive 2003/15/EC of the European parliament and the council of 27 February 2003 amending council directive 76/768/EEC on the approximation of the laws of the members states relating to cosmetic products. *Off J Eur Union L66*:26–35

- EU (2008) Council regulation of 30 May 2008 laying down test methods pursuant to regulation (EC) no. 1907/2006 of the European parliament and of the council on the registration, evaluation, authorisation and restriction of chemicals (REACH). Off J Eur Union L142:1–739
- Fabjan E, Hulzebos E, Mennes W, Piersma AH (2006) A category approach for reproductive effects of phthalates. *Crit Rev Toxicol* 36:695–726
- Farkas D, Shader RI, von Moltke LL, Greenblatt DJ (2008) Mechanisms and consequences of drug-drug interactions. In: Gad SC (ed) *Preclinical development handbook: ADME and biopharmaceutical properties*, chapter 25. Wiley, New York, pp 879–918
- Fawcett DW, Leak LV, Heidger PM Jr (1970) Electron microscopic observations on the structural components of the blood-testis barrier. *J Reprod Fert Suppl* 10:105–122
- Felter SP, Robinson MK, Basketter DA, Gerberick GF (2002) A review of the scientific basis for uncertainty factors for use in quantitative risk assessment for the induction of allergic contact dermatitis. *Contact Derm* 47(5):257–266
- Felter SP, Ryan CA, Basketter DA, Gilmour NJ, Gerberick GF (2003) Application of the risk assessment paradigm to the induction of allergic contact dermatitis. *Regul Toxicol Pharmacol* 37(1):1–10
- Felter SP, Lane RW, Latulippe ME, Llewellyn GC, Olin SS, Scimeca JA, Trautman TD (2009) Refining the threshold of toxicological concern (TTC) for risk prioritization of trace chemicals in food. *Food Chem Toxicol* 47:2236–2245
- Ferrari T, Gini G (2010) An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts. *Chem Cent J* 4(suppl 1):S2. <http://www.journal.chemistrycentral.com/content/4/S1/S2>
- Fielden MR, Brennan R, Gollub J (2007) A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. *Toxicol Sci* 99:90–100
- Fielden MR, Nie A, McMillian M, Elangbam CS, Trela BA, Yang Y, Dunn RT, Dragan Y, Fransson-Stephen R, Bogdanffy M, Adams SP, Foster WR, Chen SJ, Rossi P, Kasper P, Jacobson-Kram D, Tatsuoka KS, Wier PJ, Gollub J, Halbert DN, Roter A, Young JK, Sina JF, Marlowe J, Martus HJ, Aubrecht J, Olaharski AJ, Roome N, Nioi P, Pardo I, Snyder R, Perry R, Lord P, Mattes W, Car BD (2008) Interlaboratory evaluation of genomic signatures for predicting carcinogenicity in the rat. *Toxicol Sci* 103:28–34
- Fjodorova N, Vracko M, Novic M, Roncaglioni A, Benfenati E (2010) New public QSAR model for carcinogenicity. *Chem Cent J* 4(suppl 1):S3
- Flint OP (1983) A micromass culture method for rat embryonic neural cells. *J Cell Sci* 61:247–262
- Flint OP, Orton TC (1984) An in vitro assay for teratogens with cultures of rat embryo midbrain and limb bud cells. *Toxicol Appl Pharmacol* 76:383–395
- Fort JD, Paul RRP (2002) Enhancing the predictive validity of frog embryo teratogenesis assay—xenopus (FETAX). *J Appl Toxicol* 22:185–191
- Fowler P, Smith K, Jong J, Jeffrey L, Kirkland D, Pfulher S, Carmichael P (in preparation) Reduction of misleading (false) positive results in mammalian genotoxicity assays. I. Choice of cell type. *Mutat Res*
- Franchini JL, Propst JT, Comer GR, Yost MJ (2007) Novel tissue engineered tubular heart tissue for in vitro pharmaceutical toxicity testing. *Microsc Microanal* 13(4):267–271
- Frankild S, Vølund A, Wahlberg JE, Andersen KE (2000) Comparison of the sensitivities of the Buehler test and the guinea pig maximization test for predictive testing of contact allergy. *Acta Derm Venereol* 80(4):256–262
- Frankos VH (1985) FDA perspectives on the use of teratology data for human risk assessment. *Fundam Appl Toxicol* 5:615–625
- Freyberger A, Weimer M, Tran HS, Ahr HJ (2009a) Assessment of a recombinant androgen receptor binding assay: initial steps towards validation. *Reprod Toxicol* 30(1):2–8
- Freyberger A, Witters H, Weimer M, Lofink W, Berckmans P, Ahr HJ (2009b) Screening for (anti)androgenic properties using a standard operation protocol based on the human stably transfected androgen sensitive PALM cell line. First steps towards validation. *Reprod Toxicol* 30(1):9–17
- Freyberger A, Wilson V, Weimer M, Tan S, Tran HS, Ahr HJ (2010) Assessment of a robust model protocol with accelerated throughput for a human recombinant full length estrogen receptor-alpha binding assay: protocol optimization and intra-laboratory assay performance as initial steps towards validation. *Reprod Toxicol* 30(1):50–59
- Galvão dos Santos G, Reinders J, Ouwehand K, Rustemeyer T, Scheper RJ, Gibbs S (2009) Progress on the development of human in vitro dendritic cell based assays for assessment of the sensitising potential of a compound. *Toxicol Appl Pharmacol* 236(3):372–382
- Garber P, Ball M, Borg N, Cecchelli R, Fenart L, Hurst RD, Lindmark T, Mabondzo A, Nilsson JE, Raub TJ, Stanimirovic D, Terasaki T, Oberg JO, Osterberg T (2005) In vitro models for the blood–brain barrier. *Toxicol In Vitro* 19(3):299–334
- Garcia-Domenech R, de Julian-Ortiz JV, Besalu E (2006) True prediction of lowest observed adverse effect levels. *Mol Divers* 10(2):159–168
- Gargas ML, Burgess RJ, Voisard DE, Cason GH, Andersen ME (1989) Partition coefficients of low-molecular-weight volatile chemicals in various liquids and tissues. *Toxicol Appl Pharmacol* 98:87–99
- Gelman A, Bois F, Jiang JM (1996) Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *J Am Stat Assoc* 91:1400–1412
- Genschow E, Spielmann H, Scholz G, Seiler A, Brown N, Piersma A, Brady M, Clemann N, Huuskonen H, Paillard F, Bremer S, Becker K (2002) The ECVAM international validation study on in vitro embryotoxicity tests: results of the definitive phase and evaluation of prediction models. European centre for the validation of alternative methods. *Altern Lab Anim* 30:151–176
- Georgopoulos PG, Roy A, Gallo MA (1994) Reconstruction of short-term multi-route exposure to volatile organic compounds using physiologically based pharmacokinetic models. *J Expo Anal Environ Epidemiol* 4:309–328
- Gerberick GF, Robinson MK, Felter SP, White IR, Basketter DA (2001) Understanding fragrance allergy using an exposure-based risk assessment approach. *Contact Derm* 45(6):333–340
- Gerberick GF, Vassallo JD, Bailey RE, Chaney JG, Morrall SW, Lepoittevin JP (2004) Development of a peptide reactivity assay for screening contact allergens. *Toxicol Sci* 81(2):332–343
- Gerberick GF, Vassallo JD, Foertsch LM, Price BB, Chaney JG, Lepoittevin JP (2007) Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach. *Toxicol Sci* 97(2):417–427
- Gerberick F, Aleksic M, Basketter D, Casati S, Karlberg AT, Kern P, Kimber I, Lepoittevin JP, Natsch A, Ovigne JM, Rovida C, Sakaguchi H, Schultz T (2008) Chemical reactivity measurement and the predictive identification of skin sensitizers. *Altern Lab Anim* 36(2):215–242
- Gerberick GF, Troutman JA, Foertsch LM, Vassallo JD, Quijano M, Dobson RL, Goebel C, Lepoittevin JP (2009) Investigation of peptide reactivity of prohaptens skin sensitizers using a peroxidase-peroxide oxidation system. *Toxicol Sci* 112(1):164–174
- Gerlowski LE, Jain RK (1983) Physiologically based pharmacokinetic modeling: principles and applications. *J Pharmacol Sci* 72:1103–1127

- Ghisari M, Bonefeld-Jorgensen EC (2009) Effects of plasticizers and their mixtures on estrogen receptor and thyroid hormone functions. *Toxicol Lett* 189:67–77
- Gibaldi M, Perrier D (1982) *Pharmacokinetics*, 2nd edn, revised and expanded ed. Marcel Dekker, New York
- Gibbs S, van de Sandt JJ, Merk HF, Lockley DJ, Pendlington RU, Pease CK (2007) Xenobiotic metabolism in human skin and 3D human skin reconstructs: a review. *Curr Drug Metab* 8(8):758–772
- Gilani SH, Alibhai Y (1990) Teratogenicity of metals to chick embryos. *J Toxicol Environ Health* 30:23–31
- Gminski R, Tang T, Mersch-Sundermann V (2010) Cytotoxicity and genotoxicity in human lung epithelial A549 cells caused by airborne volatile organic compounds emitted from pine wood and oriented strand boards. *Toxicol Lett* 196(1):33–41
- Gold LS, Manley NB, Slone TH, Garfinkel GB, Ames BN, Rohrbach L, Stern BR, Chow K (1995) Sixth plot of the carcinogenic potency database: results of animal bioassays published in the general literature 1989–1990 and by the national toxicology program 1990–1993. *Environ Health Perspect* 103(suppl 8):3–122
- Gómez-Lechón MJ, Donato MT, Castell JV, Jover R (2003) Human hepatocytes as a tool for studying toxicity and drug metabolism. *Curr Drug Metab* 4(4):292–312
- Gómez-Lechón MJ, Castell JV, Donato MT (2008) An update on metabolism studies using human hepatocytes in primary culture. *Expert Opin Drug Metab Toxicol* 4(7):837–854
- Greim H, Arand M, Autrup H, Bolt HM, Bridges J, Dybing E, Glomot R, Foa V, Schulte-Hermann R (2006) Toxicological comments to the discussion about REACH. *Arch Toxicol* 80(3):121–124
- Griem P, Goebel C, Scheffler H (2003) Proposal for a risk assessment methodology for skin sensitisation based on sensitisation potency data. *Regul Toxicol Pharmacol* 38(3):269–290
- Griffiths CE, Dearman RJ, Cumberbatch M, Kimber I (2005) Cytokines and Langerhans cell mobilisation in mouse and man. *Cytokine* 32(2):67–70
- Grindon C, Combes R, Cronin MT, Roberts DW, Garrod JF (2006) Integrated testing strategies for use in the EU REACH system. *Altern Lab Anim* 34(4):407–427
- Grindon C, Combes R, Cronin MT, Roberts DW, Garrod JF (2008) An integrated decision-tree testing strategy for repeat dose toxicity with respect to the requirements of the EU REACH legislation. *Altern Lab Anim* 36(suppl 1):93–101
- Gubbels-van Hal WM, Blaauboer BJ, Barentsen HM, Hoitink MA, Meerts IA, van der Hoeven JC (2005) An alternative approach for the safety evaluation of new and existing chemicals, an exercise in integrated testing. *Regul Toxicol Pharmacol* 42(3):284–295
- Guguen-Guillouzo C, Corlu A, Guillouzo A (2010) Stem cell-derived hepatocytes and their use in toxicology. *Toxicology* 270(1):3–9
- Guironnet G, Dalbiez-Gauthier C, Rousset F, Schmitt D, Peguet-Navarro J (2000) In vitro human T cell sensitisation to haptens by monocyte-derived dendritic cells. *Toxicol In Vitro* 14(6):517–522
- Gülden M, Seibert H (2003) In vitro-in vivo extrapolation: estimation of human serum concentrations of chemicals equivalent cytotoxic concentrations in vitro. *Toxicology* 189:211–222
- Gülden M, Mörchel S, Seibert H (2001) Factors influencing nominal effective concentrations of chemical compounds in vitro: cell concentration. *Toxicol In Vitro* 15:233–243
- Guo L, Lobenhofer EK, Wang C, Shippy R, Harris SC, Zhang L, Mei N, Chen T, Herman D, Goodsaid FM, Hurban P, Philips KL, Deng QuJ, Sun YA X, Tong W, Dragon WT, Shi L (2006) Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nat Biotechnol* 24:1162–1169
- Guyton KZ, Kyle AD, Aubrecht J, Cogliano VJ, Eastmond DA, Jackson M, Keshava N, Sandy MS, Sonawane B, Zhang L, Waters MD, Smith MT (2009) Improving prediction of chemical carcinogenicity by considering multiple mechanisms and applying toxicogenomic approaches. *Mutat Res* 681:230–240
- Hah SS (2009) Recent advances in biomedical applications of accelerator mass spectrometry. *J Biomed Sci* 16:54. doi:10.1186/1423-0127-16-54
- Hallier-Vanuxeem D, Prieto P, Culot M, Diallo H, Landry C, Tati H, Cecchelli R (2009) New strategy for altering central nervous system toxicity: integration of blood brain barrier toxicity and permeability in neurotoxicity assessment. *Toxicol In Vitro* 23:447–453
- Harry GJ, Billingsley M, Bruinink A, Campbell IL, Classen W, Dorman DC, Galli C, Ray D, Smith RA, Tilson HA (1998) In vitro techniques for the assessment of neurotoxicity. *Environ Health Perspect* 106(suppl 1):131–158
- Hartig PC, Cardon MC, Blystone CR, Gray LE Jr, Wilson VS (2008) High throughput adjustable 96-well plate assay for androgen receptor binding: a practical approach for EDC screening using the chimpanzee AR. *Toxicol Lett* 181:126–131
- Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Hoffmann S, Roi AJ, Prieto P, Sabbioni E, Scott L, Worth A, Zuang V (2004) A modular approach to the ECVAM principles on test validity. *Altern Lab Anim* 32:467–472
- Hastwell PW, Chai LL, Roberts KJ, Webster TW, Harvey JS, Rees RW, Walmsley RM (2006) High-specificity and high-sensitivity genotoxicity assessment in a human cell line: validation of the GreenScreen HC GADD45a-GFP genotoxicity assay. *Mutat Res* 607:160–175
- Hattis D, Chu M, Rahmioglu N, Goble R, Verma P, Hartman K, Kozlak M (2008) A preliminary operational classification system for nonmutagenic modes of action for carcinogenesis. *Critical Rev Toxicol* 10:97–138
- Hauser C, Katz SI (1990) Generation and characterization of T-helper cells by primary in vitro sensitisation using Langerhans cells. *Immunol Rev* 117:67–84
- Hawksworth GM (2005) Isolation and culture of human renal cortical cells with characteristics of proximal tubules. *Methods Mol Med* 107:283–290
- Hayden P, Letasiova S, Kandarova H, Jackson G, Klausner M (2010) Development of a vapour cup dosing method for evaluation of chemical toxicity in the EpiAirway™ organotypic in vitro human airway model: an animal alternative model for inhalation toxicity evaluation. *ALTEX* 27(suppl 2/10):52
- Helma C (2006) Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and salmonella mutagenicity. *Mol Divers* 10:147–158
- Henkens T, Papeleu P, Elaut G, Vinken M, Rogiers V, Vanhaecke T (2007) Trichostatin A, a critical factor in maintaining the functional differentiation of primary cultured rat hepatocytes. *Toxicol Appl Pharmacol* 218(1):64–71
- Heringa MB, Schreurs RHMM, Busser F, van der Saag PT, van der Burg B, Hermens JL (2004) Towards more useful in vitro toxicity data with measured free concentrations. *Environ Sci Technol* 38:6263–6270
- Hernández LG, van Steeg H, Luijten M, van Benthem J (2009) Mechanisms of non-genotoxic carcinogens and importance of a weight of evidence approach. *Mutat Res* 682:94–109
- Hernández LG, Slob W, van Steeg H, van Benthem J (submitted) Carcinogen risk assessment using in vivo genotoxicity data and the benchmark dose approach. *Environ Mol Mutagen*
- Hewitt NJ, Lechón MJ, Houston JB, Hallifax D, Brown HS, Maurel P, Kenna JG, Gustavsson L, Lohmann C, Skonberg C, Guillouzo A, Tuschl G, Li AP, LeCluyse E, Groothuis GM, Hengstler JG (2007a) Primary hepatocytes: current understanding of the regulation of metabolic enzymes and transporter proteins, and

- pharmaceutical practice for the use of hepatocytes in metabolism, enzyme induction, transporter, clearance, and hepatotoxicity studies. *Drug Metab Rev* 39(1):159–234
- Hewitt M, Madden JC, Rowe PH, Cronin MT (2007b) Structure-based modelling in reproductive toxicology: (Q)SARs for the placental barrier. *SAR QSAR Environ Res* 18:57–76
- Hewitt M, Ellison CM, Enoch SJ, Madden JC, Cronin MTD (2010) Integrating (Q)SAR models, expert systems and read-across approaches for the prediction of developmental toxicity. *Reprod Toxicol* 30(1):147–160
- Hirota M, Suzuki M, Hagino S, Kagatani S, Sasaki Y, Aiba S, Itagaki H (2009) Modification of cell-surface thiols elicits the activation of human monocytic cell line THP-1: possible involvement in the effect of haptens, 2–4 dinitrochlorobenzene and nickel sulfate. *J Toxicol Sci* 34(2):139–150
- Hooyberghs J, Schoeters E, Lambrechts N, Nelissen I, Witters H, Schoeters G, Van Den Heuvel R (2008) A cell-based in vitro alternative to identify skin sensitizers by gene expression. *Toxicol Appl Pharmacol* 231(1):103–111
- Houck KA, Kavlock RJ (2008) Understanding mechanisms of toxicity: insights from drug discovery research. *Toxicol Appl Pharmacol* 227(2):163–178
- House RV, Thomas PT (1995) In vitro induction of cytotoxic T lymphocytes. In: Bursleson GR, Dean JH, Munson AE (eds) *Methods in immunotoxicology*. Wiley-Liss, New York, pp 159–171
- Houston JB, Galetin A (2008) Methods for predicting in vivo pharmacokinetics using data from in vitro assays. *Curr Drug Metab* 9(9):940–951
- Hu T, Gibson DP, Carr GJ, Torontali SM, Tiesman JP, Chaney JG, Aardema MJ (2004) Identification of a gene expression profile that discriminates indirect-acting genotoxins from direct-acting genotoxins. *Mutat Res* 549:5–27
- Hu T, Kaluzhny Y, Mun GC, Barnett B, Karetsky V, Wilt N, Klausner M, Curren R, Aardema MJ (2009) Intralaboratory and interlaboratory evaluation of the EpiDerm™ 3D human reconstructed skin micronucleus (RSMN) assay. *Mutat Res* 673:100–108
- Hu T, Khambatta ZS, Hayden PJ, Bolmarcich J, Binder RL, Robinson MK, Carr GJ, Tiesman JP, Jarrold BB, Osborne R, Reichling TD, Nemeth ST, Aardema MJ (2010) Xenobiotic metabolism gene expression in the EpiDerm in vitro 3D human epidermis model compared to human skin. *Toxicol In Vitro* 24:1450–1463
- Huang S (2009) Standardization and pre-validation of MucilAir: a novel in vitro cell model of the human airway epithelium for testing acute and chronic effects of chemical compounds. 3R research foundation Switzerland 3R-project 106-07. http://www.forschung3r.ch/en/projects/pr_106_07.html. Accessed 15 Jan 2011
- Hurt ME, Cappon GD, Browning A (2003) Proposal for a tiered approach to developmental toxicity testing for veterinary pharmaceutical products for food-producing animals. *Food Chem Toxicol* 41:611–619
- Hwan Sung J, Esch MB, Shuler ML (2010) Integration of in silico and in vitro platforms for pharmacokinetic-pharmacodynamic modeling. *Expert Opin Drug Metab Toxicol* 6:1–19
- ILSI HESI ACT (2001) ILSI HESI alternatives to carcinogenicity testing project. *Toxicol Pathol* 29(supplement):1–351
- Isfort RJ, Kerckaert GA, LeBoeuf RA (1996) Comparison of the standard and reduced pH Syrian hamster embryo (SHE) cell in vitro transformation assays in predicting the carcinogenic potential of chemicals. *Mutat Res* 356:11–63
- Ishihara A, Rahman FB, Leelawatwattana L, Prapunpoj P, Yamauchi K (2009) In vitro thyroid hormone-disrupting activity in effluents and surface waters in Thailand. *Environ Toxicol Chem* 28:586–594
- Jäckh C, Blatz V, Guth K, Reisinger K, Fabian E, van Ravenzwaay B, Landsiedel R (2010) Enzyme activities for xenobiotic metabolism in human reconstructed skin models. *ALTEX* 27(suppl 2/10):60
- Jagger C, Tate M, Cahill PA, Hughes C, Knight AW, Billinton N, Walmsley RM (2009) Assessment of the genotoxicity of S9-generated metabolites using the GreenScreen HC GADD45a-GFP assay. *Mutagenesis* 24:35–50
- Jamei M, Dickinson GL, Rostami-Hodjegan A (2009) A framework for assessing interindividual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology and genetics: a tale of ‘bottom-up’ vs ‘top-down’ recognition of covariates. *Drug Metab Pharmacokinet* 24:53–75
- Jaworska J, Nikolova-Jeliazkova N (2007) How can structural similarity analysis help in category formation. *SAR QSAR Environ Res* 18:195–207
- JECFA (1996) A procedure for the safety evaluation of flavouring agents. WHO food additive series 35. Annex 5. World Health Organization, Geneva
- JECFA (1997) Safety evaluation of certain food additives and contaminants. Forty-ninth report of the joint FAO/WHO expert committee on food additives. WHO food additive series 40. World Health Organization, Geneva
- Jelinek R (1977) The chick embryotoxicity screening test (CHEST). In: Neubert D, Merker HJ, Kwasigroch TE (eds) *Methods in prenatal toxicology*. Thieme, Stuttgart, pp 381–386
- Jelinek R, Marhan O (1994) Validation of the chick embryotoxicity screening test (CHEST). A comparative study. *Funct Dev Morphol* 4:317–323
- Jelinek R, Peterka M, Rychter Z (1985) Chick embryotoxicity screening test—130 substances tested. *Indian J Exp Biol* 23: 588–595
- Jennings P (2010) The role of the Nrf2 pathway in nephrotoxicity. *ALTEX* 27(suppl 2/10):60–61
- Jennings P, Aydin S, Bennett J, McBride R, Weiland C, Tuite N, Gruber LN, Perco P, Gaora PO, Ellinger-Ziegelbauer H, Ahr HJ, Van Kooten C, Daha MR, Prieto P, Ryan MP, Pfaller W, McMorro T (2009) Inter-laboratory comparison of human renal proximal tubule (HK-2) transcriptome alterations due to Cyclosporine A exposure and medium exhaustion. *Toxicol In Vitro* 23(3):486–499
- Jensen GE, Niemela JR, Wedebye EB, Nikolov NG (2008) QSAR models for reproductive toxicity and endocrine disruption in regulatory use—a preliminary investigation. *SAR QSAR Environ Res* 19:631–641
- Jiang Y, Gerhold DL, Holder DJ, Figueroa DJ, Bailey WJ, Guan P, Skopek TR, Sistare FR, Sina JF (2007) Diagnosis of drug-induced renal tubular toxicity using global gene expression profiles. *J Transl Med* 5:47
- Jones HM, Gardner IB, Watson KJ (2009) Modelling and PBPK simulation in drug discovery. *AAPS J* 11:155–166
- Jonker MJ, Bruning O, van Iterson M, Schaap MM, van der Hoeven TV, Vrieling H, Beems RB, de Vries A, van Steeg H, Breit TM, Luijten M (2009) Finding transcriptomics biomarkers for in vivo identification of (non-)genotoxic carcinogens using wild-type and Xpa/p53 mutant mouse models. *Carcinogenesis* 30:1805–1812
- Jowsey IR, Basketter DA, Westmoreland C, Kimber I (2006) A future approach to measuring relative skin sensitising potency: a proposal. *J Appl Toxicol* 26:341–350
- Kane KL, Ashton FA, Schmitz JL, Folds JD (1996) Determination of natural killer cell function by flow cytometry. *Clin Diagn Lab Immunol* 3(3):295–300
- Kapitulnik J, Pelkonen O, Gundert-Remy U, Dahl SG, Boobis AR (2009) Effects of pharmaceuticals and other active chemicals at biological targets: mechanisms, interactions, and integration into PB-PK/PD models. *Expert Opin Therap Targets* 13:867–887

- Karlberg AT, Basketter D, Goossens A, Lepoittevin JP (1999) Regulatory classification of substances oxidized to skin sensitizers by exposure to the air. *Contact Derm* 40(4):183–188
- Kasting G, Farahmand S, Nitsche J, Kern P, Gerberick F (2010) Epidermal bioavailability of volatile compounds. Application to fragrance disposition and skin sensitization risk assessment. *Perfumer Flavorist* 35
- Kawakami T, Ishimura R, Nohara K, Takeda K, Tohyama C, Ohsako S (2006) Differential susceptibilities of Holtzman and Sprague–Dawley rats to fetal death and placental dysfunction induced by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) despite the identical primary structure of the aryl hydrocarbon receptor. *Toxicol Appl Pharmacol* 212:224–236
- Keller D, Krauledat M, Scheel J (2009) Feasibility study to support a threshold of sensitization concern concept in risk assessment based on human data. *Arch Toxicol* 83(12):1049–1060
- Kemper FH, Luepke NP (1986) Toxicity testing by the hens egg test (het). *Food Chem Toxicol* 24:647–648
- Kettenhofen R, Bohlen H (2008) Preclinical assessment of cardiac toxicity. *Drug Discov Today* 13(15–16):702–707
- Kienhuis AS, Wortelboer HM, Maas WJ, van Herwijnen M, Kleinjans JC, van Delft JH, Stierum RH (2007) A sandwich-cultured rat hepatocyte system with increased metabolic competence evaluated by gene expression profiling. *Toxicol In Vitro* 21(5):892–901
- Kier LD, Neft R, Tang L, Suizu R, Cook T, Onsurez K, Tiegler K, Sakai Y, Ortiz M, Nolan T, Sankar U, Li AP (2004) Applications of microarrays with toxicologically relevant genes (tox genes) for the evaluation of chemical toxicants in Sprague–Dawley rats in vivo and human hepatocytes in vitro. *Mutat Res* 549(1–2):101–113
- Kikkawa R, Yamamoto T, Fukushima T, Yamada H, Horii I (2005) Investigation of a hepatotoxicity screening system in primary cell cultures: “what biomarkers would need to be addressed to estimate toxicity in conventional and new approaches?”. *J Toxicol Sci* 30(1):61–72
- Kikkawa R, Fujikawa M, Yamamoto T, Hamada Y, Yamada H, Horii I (2006) In vivo hepatotoxicity study of rats in comparison with in vitro hepatotoxicity screening system. *J Toxicol Sci* 31(1):23–34
- Kim MS, Lee W, Kim YC, Park JK (2008) Microvalve-assisted patterning platform for measuring cellular dynamics based on 3D cell culture. *Biotechnol Bioeng* 101(5):1005–1013
- Kimber I, Basketter DA (1992) The murine local lymph node assay: a commentary on collaborative studies and new directions. *Food Chem Toxicol* 30(2):165–169
- Kimber I, Cumberbatch M, Dearman RJ, Bhushan M, Griffiths CE (2000) Cytokines and chemokines in the initiation and regulation of epidermal Langerhans cell mobilization. *Br J Dermatol* 142(3):401–412
- Kimber I, Dearman RJ, Betts CJ, Gerberick GF, Ryan CA, Kern PS, Patlewicz GY, Basketter DA (2006) The local lymph node assay and skin sensitisation: a cut-down screen to reduce animal requirements? *Contact Derm* 54(4):181–185
- Kirkland D, Fowler P (2010) Further analysis of Ames-negative rodent carcinogens that are only genotoxic in mammalian cells in vitro at concentrations exceeding 1 mM, including retesting of compounds of concern. *Mutagenesis* 25:539–553
- Kirkland D, Speit G (2008) Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens III. Appropriate follow-up testing in vivo. *Mutat Res* 654:114–132
- Kirkland D, Aardema M, Henderson L, Müller L (2005) Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity. *Mutat Res* 584:1–256
- Kirkland D, Aardema M, Müller L, Makoto H (2006) Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens II. Further analysis of mammalian cell results, relative predictivity and tumour profiles. *Mutat Res* 608:29–42
- Kirkland D, Pfuhrer 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. *Mutat Res* 628:31–55
- Kirkland D, Kasper P, Müller L, Corvi R, Speit G (2008) Recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: a follow-up to an ECVAM workshop. *Mutat Res* 653:99–108
- Klaunig JE, Shi Y (2009) Assessment of gap junctional intercellular communication. *Curr Protoc Toxicol*, chapter 2, unit 2.17
- Klaunig JE, Kamendulis LM, Hoocevar BA (2010) Oxidative stress and oxidative damage in carcinogenesis. *Toxicol Pathol* 38:96–109
- Klopman G, Rosenkranz H (1994) Approaches to SAR in carcinogenesis and mutagenesis. Prediction of carcinogenicity/mutagenicity using MULTI-CASE. *Mutat Res* 305:33–46
- Klotz DM, Ladlie BL, Vonier PM, McLachlan JA, Arnold SF (1997) *o,p'*-DDT and its metabolites inhibit progesterone-dependent responses in yeast and human cells. *Mol Cell Endocrinol* 129:63–71
- Knight A, Jonker MJ, Bruning J (2005) Which drugs cause cancer? *BMJ USA* 5:477–478
- Knight A, Bailey J, Balcombe J (2006) Animal carcinogenicity studies: implications for the REACH system. *Altern Lab Anim* 34(suppl 1):139–147
- Kotwani A (1998) Use of chick embryo in screening for teratogenicity. *Indian J Physiol Pharmacol* 42:189–204
- Kramer NI (2010) Measuring, modeling and increasing the free concentration of test chemicals in cell assays. PhD thesis, Utrecht University
- Kramer NI, Hermens JLM, Schirmer K (2009) The influence of modes of action and physicochemical properties of chemicals on the correlation between in vitro and acute fish toxicity data. *Toxicol In Vitro* 23:1372–1379
- Kreiling R, Hollnagel HM, Hareng L, Eigler D, Lee MS, Griem P, Dreesen B, Kleber M, Albrecht A, Garcia C, Wendel A (2008) Comparison of the skin sensitizing potential of unsaturated compounds and assessed by the murine local lymph node assay (LLNA) and the guinea pig maximization test (GPMT). *Food Chem Toxicol* 46(6):1896–1904
- Krewski D, Acosta D Jr, Andersen M, Anderson H, Bailar JC III, Boekelheide K, Brent R, Charnley G, Cheung VG, Green S Jr, Kelsey KT, Kerkvliet NI, Li AA, McCray L, Meyer O, Patterson RD, Pennie W, Scala RA, Solomon GM, Stephens M, Yager J, Zeise L (2010) Toxicity testing in the 21st century: a vision and a strategy. *J Toxicol Environ Health B Crit Rev* 13:51–138
- Kroes R, Koziarowski G (2002) Threshold of toxicological concern (TTC) in food safety assessment. *Toxicol Lett* 127(1–3):43–46
- Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothorst F, Vos JG, Würtzen G (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 concept in risk assessment. *Toxicol Sci* 86(2):226–230
- Kroes R, Renwick AG, Feron V, Galli CL, Gibney M, Greim H, Guy RH, Lhuguenot JC, van de Sandt JJ (2007) Application of the

- threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients. *Food Chem Toxicol* 45:2533–2562
- Krüse J, Golden D, Wilkinson S, Williams F, Kezic S, Corish J (2007) Analysis, interpretation, and extrapolation of dermal permeation data using diffusion-based mathematical models. *J Pharmacol Sci* 96:682–703
- Kucera P, Burnand MB (1987) Routine teratogenicity test that uses chick embryos in vitro. *Teratog Carcinog Mutagen* 7:427–447
- Kulkarni SA, Zhu J, Blechinger S (2005) In silico techniques for the study and prediction of xenobiotic metabolism: a review. *Xenobiotica* 35(10–11):955–973
- La SG, Farini D, De FM (2010) Estrogenic in vitro assay on mouse embryonic Leydig cells. *Int J Dev Biol* 54:717–722
- Lambrechts N, Vanheel H, Hooyberghs J, De Boever P, Witters H, Van Den Heuvel R, Van Tendeloo V, Nelissen I, Schoeters G (2010) Gene markers in dendritic cells unravel pieces of the skin sensitisation puzzle. *Toxicol Lett* 196(2):95–103
- Lambrot R, Muczynski V, Lecureuil C, Angenard G, Coffigny H, Pairault C, Moison D, Frydman R, Habert R, Rouiller-Fabre V (2009) Phthalates impair germ cell development in the human fetal testis in vitro without change in testosterone production. *Environ Health Perspect* 117:32–37
- Landsiedel R, Fabian E, Gamer A, Kolle S, Ma-Hock L, Schulz M, Wiench K, Wohlleben W, van Ravenzwaay B (2010) The use of alternative methods for toxicity testing of nanomaterials. *ALTEX* 27 (suppl 2/10):78
- Langezaal I, Hoffmann S, Hartung T, Coecke S (2002) Evaluation and prevalidation of an immunotoxicity test based on human whole-blood cytokine release. *Altern Lab Anim* 30(6):581–595
- Lankveld DP, Van Loveren H, Baken KA, Vandebriel RJ (2010) In vitro testing for direct immunotoxicity: state-of-the-art, chapter 26. In: Dietert RR (ed) *Immunotoxicity testing: methods and protocols methods in molecular biology*. Humana Press Inc., Totowa, pp 401–423
- Lappin G, Garner RC (2005) The use of accelerator mass spectrometry to obtain early human ADME/PK data. *Expert Opin Drug Metab Toxicol* 1(1):23–31
- Lazzari G, Tessaro I, Crotti G, Galli C, Hoffmann S, Bremer S, Pellizzer C (2008) Development of an in vitro test battery for assessing chemical effects on bovine germ cells under the ReProTect umbrella. *Toxicol Appl Pharmacol* 233:360–370
- Le Fevre AC, Boitier E, Marchandean JP, Sarasin A, Thybaud V (2007) Characterization of DNA reactive and non-DNA reactive anticancer drugs by gene expression profiling. *Mutat Res* 619:16–29
- Lee PJ, Hung PJ, Lee LP (2007) An artificial liver sinusoid with a microfluidic endothelial-like barrier for primary hepatocyte culture. *Biotechnol Bioeng* 97(5):1340–1346
- Legendre A, Froment P, Desmots S, Lecomte A, Habert R, Lemazurier E (2010) An engineered 3D blood-testis barrier model for the assessment of reproductive toxicity potential. *Biomaterials* 31:4492–4505
- Lemeire K, Van Merris V, Cortvrindt R (2007) The antibiotic streptomycin assessed in a battery of in vitro tests for reproductive toxicology. *Toxicol In Vitro* 21:1348–1353
- Lewis DF, Ito Y (2008) Human cytochromes P450 in the metabolism of drugs: new molecular models of enzyme-substrate interactions. *Experts Opin Drug Metab Toxicol* 4(9):1181–1186
- Li AP (2008) In vitro evaluation of metabolic drug-drug interactions: scientific concepts and practical considerations. In: Gad SC (ed) *Preclinical development handbook: ADME and biopharmaceutical properties*, chapter 24. Wiley, New York, pp 853–878
- Li Z, Chan C (2009) Systems biology for identifying liver toxicity pathways. *BMC Proc* 3(suppl 2):S2
- Li W, Lam M, Choy D, Birkeland A, Sullivan ME, Post JM (2006) Human primary renal cells as a model for toxicity assessment of chemo-therapeutic drugs. *Toxicol In Vitro* 20(5):669–676
- Li HH, Aubrecht J, Fornace AJ Jr (2007) Toxicogenomics: overview and potential applications for the study of non-covalent DNA interacting chemicals. *Mutat Res* 623:98–108
- Liao KH, Tan YM, Clewell HJ (2007) Development of a screening approach to interpret human biomonitoring data on volatile organic compounds: reverse dosimetry on biomonitoring data for trichloroethylene. *Risk Anal* 27:1223–1236
- Lima BS, van der Laan JW (2000) Mechanisms of nongenotoxic carcinogenesis and assessment of the human hazard. *Regul Toxicol Pharmacol* 32:135–143
- Lippman M, Bolan G, Huff K (1976) The effects of estrogens and antiestrogens on hormone-responsive human breast cancer in long-term tissue culture. *Cancer Res* 36:4595–4601
- Lo PK, Sukumar S (2008) Epigenomics and breast cancer. *Pharmacogenomics* 9:1879–1902
- Loeb LA, Harris CC (2008) Advances in chemical carcinogenesis: a historical review and prospective. *Cancer Res* 68:6863–6872
- Loizou GD, Spendiff M, Barton HA, Bessems J, Bois FY, d'Yvoire MB, Buist H, Clewell HJ III, Gundert-Remy U, Goerlitz G, Meek B, Schmitt W (2008) Development of good modelling practice for physiologically based pharmacokinetic models for use in risk assessment: the first steps. *Regul Toxicol Pharmacol* 50:400–411
- Long M, Laier P, Vinggaard AM, Andersen HR, Lynggaard J, Bonfeld-Jorgensen EC (2003) Effects of currently used pesticides in the AhR-CALUX assay: comparison between the human TV101L and the rat H4IIE cell line. *Toxicology* 194:77–93
- López-Lázaro M (2010) A new view of carcinogenesis and an alternative approach to cancer therapy. *Mol Med* 16:144–153
- Luciano AM, Franciosi F, Lodde V, Corbani D, Lazzari G, Crotti G, Galli C, Pellizzer C, Bremer S, Weimer M, Modena SC (2010) Transferability and inter-laboratory variability assessment of the in vitro bovine oocyte maturation (IVM) test within ReProTect. *Reprod Toxicol* 30(1):81–88
- Luijten M, van Beelen VA, Verhoef A, Renkens MF, van Herwijnen MH, Westerman A, van Schooten FJ, Pennings JL, Piersma AH (2010) Transcriptomics analysis of retinoic acid embryotoxicity in rat postimplantation whole embryo culture. *Reprod Toxicol* 30:333–340
- Lyons MA, Yang RSH, Mayeno AN, Reisfeld B (2008) Computational toxicology of chloroform: reverse dosimetry using bayesian inference, markov chain monte carlo simulation, and human biomonitoring data. *Environmental and health Perspectives* 116:1040–1046
- Ma B, Zhang G, Qin J, Lin B (2009) Characterization of drug metabolites and cytotoxicity assay simultaneously using an integrated microfluidic device. *Lab Chip* 9(2):232–238
- Madden JC (2010) In silico approaches for predicting ADME properties. In: Puzyn T, Leszczynski J, Cronin MTD (eds) *Recent advances in QSAR studies—methods and applications*. Springer, Dordrecht, pp 283–304
- Magnusson B, Kligman AM (1970) Allergic contact dermatitis in the guinea pig: identification of contact allergens. Charles C. Thomas, Springfield
- MAQC Consortium (2006) The microarray quality control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24:1151–1161
- Marchant C, Fisk L, Note R, Patel M, Surez D (2009) An expert system approach to the assessment of hepatotoxic potential. *Chem Biodivers* 6(11):2107–2114
- Martin MT, Judson RS, Reif DM, Kavlock RJ, Dix DJ (2009) Profiling chemicals based on chronic toxicity results from the US EPA ToxRef Database. *Environ Health Perspect* 117:392–399
- Martin SF, Esser PR, Schmucker S, Dietz L, Naisbitt DJ, Park BK, Vocanson M, Nicolas JF, Keller M, Pichler WJ, Peiser M, Luch A, Wanner R, Maggi E, Cavani A, Rustemeyer T, Richter A,

- Thierse HJ, Sallusto F (2010) T cell recognition of chemicals, protein allergens and drugs: towards the development of in vitro assays. *Cell Mol Life Sci* 67(24):4171–4184
- Martínez-Jiménez CP, Castell JV, Gómez-Lechón MJ, Jover R (2006) Transcriptional activation of CYP2C9, CYP1A1, and CYP1A2 by hepatocyte nuclear factor 4alpha requires coactivators peroxisomal proliferator activated receptor-gamma coactivator 1alpha and steroid receptor coactivator 1. *Mol Pharmacol* 70(5):1681–1692
- Marx J (2003) Building better mouse models for studying cancer. *Science* 299:1972–1975
- Marx-Stoelting P, Adriaens E, Ahr HJ, Bremer S, Garthoff B, Gelbke HP, Piersma A, Pellizzer C, Reuter U, Rogiers V, Schenk B, Schwengberg S, Seiler A, Spielmann H, Steemans M, Stedman DB, Vanparys P, Vericat JA, Verwei M, van der Water F, Weimer M, Schwarz M (2009) A review of the implementation of the embryonic stem cell test (EST). The report and recommendations of an ECVAM/ReProTect workshop. *Altern Lab Anim* 37:313–328
- Mathijs K, Brauers KJ, Jennen DGJ, Lizarraga D, Kleinjans JCS, van Delft JHM (2010) Gene expression profiling in primary mouse hepatocytes discriminates true from false-positive genotoxic compounds. *Mutagenesis* 25:561–568
- Matthews EJ, Spalding JW, Tennant RW (1993) Transformation of BALB/c-3T3 cells: V. Transformation responses of 168 chemicals compared with mutagenicity in salmonella and carcinogenicity in rodent bioassays. *Environ Health Perspect* 101(suppl 2):347–482
- Matthews EJ, Kruhlak NL, Benz RD, Contrera JF (2004a) Assessment of the health effects of chemicals in humans: I. QSAR estimation of the maximum recommended therapeutic dose (MRTD) and no effect level (NOEL) of organic chemicals based on clinical trial data. *Curr Drug Discov Technol* 1(1):61–76
- Matthews EJ, Kruhlak NL, Weaver JL, Benz RD, Contrera JF (2004b) Assessment of the health effects of chemicals in humans: II. Construction of an adverse effects database for QSAR modeling. *Curr Drug Discov Technol* 1(4):243–254
- Matthews EJ, Kruhlak NL, Benz RD, Contrera JF, Marchant CA, Yang C (2008) Combined use of MC4PC, MDL-QSAR, BioEpisteme, Leadscape PDM, and Derek for Windows Software to achieve high-performance, high-confidence, mode of action-based predictions of chemical carcinogenesis in rodents. *Toxicol Mech Methods* 18:189–206
- Maunz A, Helma C (2008) Prediction of chemical toxicity with local support vector regression and activity-specific kernels. *SAR QSAR Environ Res* 19(5–6):413–431
- Maurici D, Aardema M, Corvi R, Kleber M, Krul C, Laurent C, Loprieno N, Pasanen M, Pfuhler S, Phillips B, Sabbioni E, Sanner T, Vanparys P (2005) Genotoxicity and mutagenicity. *Altern Lab Anim* 33(suppl 1):117–130
- Maxwell G, Mackay C (2008) Application of a systems biology approach to skin allergy risk assessment. *Altern Lab Anim* 36(5):521–556
- Mazzatorta P, Tran L-A, Schilter B, Grigorov M (2007) Integration of structure activity relationship and artificial intelligence systems to improve in silico prediction of Ames test mutagenicity. *J Chem Inf Model* 47(1):34–38
- Mazzatorta P, Estevez MD, Coulet M, Schilter B (2008) Modeling oral rat chronic toxicity. *J Chem Inf Model* 48(10):1949–1954
- McFadden JP, Basketter DA (2000) Contact allergy, irritancy and danger. *Contact Derm* 42(3):123–127
- McKim JM, Keller DJ, Gorski JR (2010) A new in vitro method for identifying chemical sensitizers combining peptide binding with ARE/EpRE-mediated gene expression in human skin cells. *Cutan Ocul Toxicol* 29(3):171–192
- Meek ME (2004) Biologically motivated computational modeling: contribution to risk assessment. *Toxicol Sci* 82:1–2
- Mehdipour AR, Hamidi M (2009) Brain drug targeting: a computational approach for overcoming blood–brain barrier. *Drug Discov Today* 14:1030–1036
- Mekenyan O, Dimitrov S, Serafimova R, Thompson E, Kotov S, Dimitrova N, Walker JD (2004) Identification of the structural requirements for mutagenicity by incorporating molecular flexibility and metabolic activation of chemicals I: TA100 model. *Chem Res Toxicol* 17:753–766
- Mekenyan O, Todorov M, Serafimova R, Stoeva S, Aptula A, Finking R, Jacob E (2007) Identifying the structural requirements for chromosomal aberration by incorporating molecular flexibility and metabolic activation of chemicals. *Chem Res Toxicol* 20:1927–1941
- Mellman I, Steinman RM (2001) Dendritic cells: specialized and regulated antigen processing machines. *Cell* 106(3):255–258
- Meneses-Lorente G, Watt A, Salim K, Gaskell SJ, Muniappa N, Lawrence J, Guest PC (2006) Identification of early proteomic markers for hepatic steatosis. *Chem Res Toxicol* 19(8):986–998
- Mielke H, Anger LT, Schug M, Hengstler JG, Stahlmann R, Gundert-Remy U (2010) A physiologically based toxicokinetic modelling approach to predict relevant concentrations for in vitro testing. *Arch Toxicol* [Epub ahead of print]
- Miller CL, Eaves CJ (2002) Long-term culture-initiating cell assays for human and murine cells. In: Klug CA, Jordan CT (eds) *Haematopoietic stem cell protocols*. Humana Press Inc., New Jersey, pp 123–141
- Molina-Molina JM, Hillenweck A, Jouanin I, Zalko D, Cravedi JP, Fernandez MF, Pillon A, Nicolas JC, Olea N, Balaguer P (2006) Steroid receptor profiling of vinclozolin and its primary metabolites. *Toxicol Appl Pharmacol* 216:44–54
- Moore P, Ogilvie J, Horridge E, Mellor IR, Clothier RH (2005) The development of an innervated epithelial barrier model using a human corneal cell line and ND7/23 sensory neurons. *Eur J Cell Biol* 84(5):581–592
- Morck TJ, Sorda G, Bechi N, Rasmussen BS, Nielsen JB, Ietta F, Rytting E, Mathiesen L, Paulesu L, Knudsen LE (2010) Placental transport and in vitro effects of Bisphenol A. *Reprod Toxicol* 30(1):131–137
- Mortishire-Smith RJ, Skiles GL, Lawrence JW, Spence S, Nicholls AW, Johnson BA, Nicholson JK (2004) Use of metabonomics to identify impaired fatty acid metabolism as the mechanism of a drug-induced toxicity. *Chem Res Toxicol* 17(2):165–173
- Mose T, Kjaerstad MB, Mathiesen L, Nielsen JB, Edelfors S, Knudsen LE (2008) Placental passage of benzoic acid, caffeine, and glyphosate in an ex vivo human perfusion system. *J Toxicol Environ Health A* 71:984–991
- Moss E, Lynagh S, Smith D, Kelly S, McDauid A, Bunton D (2010) A novel system for the investigation of microvascular dysfunction including vascular permeability and flow-mediated dilatation in pressurised human arteries. *J Pharmacol Toxicol Methods* 62(1):40–46
- Mostrag-Szlichtyng A, Worth A (2010) Review of QSAR models and software tools for predicting biokinetic properties. EUR 24377 EN. European Commission, Luxembourg
- Müller L, Mauthe RJ, Riley CM, Andino MM, Antonis DD, Beels C, DeGeorge J, De Knaep AG, Ellison D, Fagerland JA, Frank R, Fritschel B, Galloway S, Harpur E, Humfrey CD, Jacks AS, Jagota N, Mackinnon J, Mohan G, Ness DK, O'Donovan MR, Smith MD, Vudathala G, Yotti L (2006) A rationale for determining, testing, and controlling specific impurities in pharmaceuticals that possess potential for genotoxicity. *Regul Toxicol Pharmacol* 44:198–211
- Mumtaz MM, Knauf LA, Reisman DJ, Peirano WB, DeRosa CT, Gombar VK, Enlein K, Carter JR, Blake BW, Huque KI,

- Ramanujam MS (1995) Assessment of effect levels of chemicals from quantitative structure–activity relationship (QSAR) models. I. Chronic lowest-observed-adverse effect level (LOAEL). *Toxicol Lett* 79(1–3):131–143
- Mun GC, Aardema MJ, Hu T, Barnett B, Kaluzhny Y, Klausner M, Karetsky V, Dahl EL, Curren RD (2009) Further development of the EpiDermTM 3D reconstructed human skin micronucleus (RSMN) assay. *Mutat Res* 673:92–99
- Munro IC, Ford RA, Kennepohl E, Sprenger JG (1996a) Thresholds of toxicological concern based on structure–activity relationships. *Drug Metab Rev* 28(1–2):209–217
- Munro IC, Ford RA, Kennepohl E, Sprenger JG (1996b) Correlation of structural class with no-observed effect levels: a proposal for establishing a threshold of concern. *Food Chem Toxicol* 34:829–867
- Munro IC, Renwick AG, Danielewska-Nikiel B (2008) The threshold of toxicological concern (TTC) in risk assessment. *Toxicol Lett* 180(2):151–156
- Muster W, Breidenbach A, Fischer H, Kirchner S, Muller L, Pahler A (2008) Computational toxicology in drug development. *Drug Discov Today* 13(7–8):303–310
- Myllynen P, Immonen E, Kumm M, Vahakangas K (2009) Developmental expression of drug metabolizing enzymes and transporter proteins in human placenta and fetal tissues. *Expert Opin Drug Metab Toxicol* 5:1483–1499
- Myren M, Mose T, Mathisen L, Knudsen LE (2007) The human placenta—an alternative to foetal exposure. *Toxicol In Vitro* 21:1332–1340
- Nagel R (2002) DarT: the embryo test with the zebrafish *danio rerio*—a general model in ecotoxicology and toxicology. *ALTEX* 19(suppl 1):38–48
- Naiki T, Nagaki M, Asano T, Kimata T, Moriwaki H (2005) Adenovirus-mediated hepatocyte nuclear factor-4 α overexpression maintains liver phenotype in cultured rat hepatocytes. *Biochem Biophys Res Commun* 335(2):496–500
- Nakai D, Kumamoto K, Sakikawa Ch, Kosaka T, Tokui T (2004) Evaluation of protein binding ratio of drugs by a micro-scale ultracentrifugation method. *J Pharmaceut Sci* 93:847–851
- Natsch A (2010) The Nrf2-Keap1-ARE toxicity pathway as a cellular sensor for skin sensitizers functional relevance and a hypothesis on innate reactions to skin sensitizers. *Toxicol Sci* 113(2):284–292
- Natsch A, Emter R (2008) Skin sensitizers induce antioxidant response element dependent genes: application to the in vitro testing of the sensitization potential of chemicals. *Toxicol Sci* 102(1):110–119
- Natsch A, Gfeller H (2008) LC–MS based characterization of the peptide reactivity of chemicals to improve the in vitro prediction of the skin sensitization potential. *Toxicol Sci* 106(2):464–478
- Natsch A, Emter R, Ellis G (2009) Filling the concept with data: integrating data from different in vitro and in silico assays on skin sensitizers to explore the battery approach for animal-free skin sensitization testing. *Toxicol Sci* 107(1):106–121
- Nestorov I (2003) Whole body pharmacokinetic models. *Clin Pharmacokinet* 42:883–908
- Nestorov I (2007) Whole-body physiologically based pharmacokinetic models. *Expert Opin Drug Metab Toxicol* 3:235–249
- Nie AY, McMillian M, Parker JB, Leone A, Bryant S, Yieh L, Bittner A, Nelson J, Carmen A, Wan J, Lord PG (2006) Predictive toxicogenomics approaches reveal underlying molecular mechanisms of nongenotoxic carcinogenicity. *Mol Carcinog* 45:914–933
- Nioi P, Pardo ID, Sherratt PJ, Snyder RD (2008) Prediction of nongenotoxic carcinogenesis in rats using changes in gene expression following acute dosing. *Chem Biol Interact* 172:206–215
- Nohynek GJ, Antignac E, Re T, Toutain H (2010) Safety assessment of personal care products/cosmetics and their ingredients. *Toxicol Appl Pharmacol* 243(2):239–259
- NRC, National Research Council (2007) Toxicity testing in the 21st century: a vision and a strategy. National Academies Press, Washington, DC
- OECD, Organisation for Economic Cooperation and Development (1981a) Test guideline on repeated dose dermal toxicity: 21/28-day study, Paris
- OECD, Organisation for Economic Cooperation and Development (1981b) Test guideline on repeated dose dermal toxicity: 90-day study, Paris
- OECD, Organisation for Economic Cooperation and Development (1983) Guideline for the testing of chemicals 415. One-generation reproduction toxicity study, Paris
- OECD, Organisation for Economic Cooperation and Development (1992) Skin sensitization guidelines for testing of chemicals no. 406, Paris
- OECD, Organisation for Economic Cooperation and Development (1995) Guideline for the testing of chemicals 421. Reproduction/developmental toxicity screening test, Paris
- OECD, Organisation for Economic Cooperation and Development (1996) Guideline for the testing of chemicals 422. Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test, Paris
- OECD, Organisation for Economic Cooperation and Development (1998a) Test guideline on repeated Dose 90-day oral toxicity study in rodents, Paris
- OECD, Organisation for Economic Cooperation and Development (1998b) Test guideline on repeated dose 90-day oral toxicity study in non-rodents, Paris
- OECD, Organisation for Economic Cooperation and Development (2001a) Guideline for the testing of chemicals 414. Prenatal development toxicity study, Paris
- OECD, Organisation for Economic Cooperation and Development (2001b) Guideline for the testing of chemicals 416. Two-generation reproduction toxicity, Paris
- OECD, Organisation for Economic Cooperation and Development (2002) The local lymph node assay. Guidelines for testing of chemicals no. 429, Paris
- OECD, Organisation for Economic Cooperation and Development (2004a) OECD series on testing and assessment number 49. Report from the expert group on (quantitative) structure–activity relationships [(Q)SARs] on the principles for the validation of (Q)SARs, Paris
- OECD, Organisation for Economic Cooperation and Development (2004b) Test guideline 428. Skin absorption: in vitro method. Available via <http://www.oecdilib.org/docserver/download/fulltext/9742801e.pdf?expires=1275558971&id=0000&acname=Dutch+Government&checksum=4396D617A72B24A90D1DE616728D8258>
- OECD, Organisation for Economic Cooperation and Development (2007a) Guidance document on the validation of (quantitative) structure–activity relationships [(Q)SAR] models, Paris
- OECD, Organisation for Economic Cooperation and Development (2007b) Guidance on grouping of chemicals, OECD environment health and safety publications, series on testing and assessment no. 80. [http://www.appl1.oecd.org/olis/2007doc.nsf/linkto/env-jm-mono\(2007\)28](http://www.appl1.oecd.org/olis/2007doc.nsf/linkto/env-jm-mono(2007)28)
- OECD, Organisation for Economic Cooperation and Development (2007c) Guideline for the testing of chemicals 426. Developmental neurotoxicity study, Paris
- OECD, Organisation for Economic Cooperation and Development (2007d) Guideline for the testing of chemicals 440. Uterotrophic bioassay in rodents. A short-term screening test for oestrogenic properties, Paris

- OECD, Organisation for Economic Cooperation and Development (2007e) Detailed review on cell transformation assays for detection of chemical carcinogens, OECD environment, health and safety publications, series on testing and assessment, no. 31. [http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=env/jm/mono\(2007\)18&doclanguage=en](http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=env/jm/mono(2007)18&doclanguage=en)
- OECD, Organisation for Economic Cooperation and Development (2008a) Test guideline on repeated Dose 28-day oral toxicity study in rodents, Paris
- OECD, Organisation for Economic Cooperation and Development (2008b) Guidance document 43: guidance document on mammalian reproductive toxicity testing and assessment, Paris
- OECD, Organisation for Economic Cooperation and Development (2009a) Test guideline on subacute inhalation toxicity: 28-day study, Paris
- OECD, Organisation for Economic Cooperation and Development (2009b). OECD guideline for the testing of chemicals 441. Hershberger bioassay in rats. A short-term screening assay for (anti)androgenic properties, Paris
- OECD, Organisation for Economic Cooperation and Development (2009c). OECD guideline for the testing of chemicals 455. The stably transfected human estrogen receptor-alpha transcriptional activation assay for detection of estrogenic agonist-activity of chemicals, Paris
- OECD, Organisation for Economic Cooperation and Development (2009d). OECD draft proposal for an extended one-generation reproductive toxicity study, Paris
- OECD, Organisation for Economic Cooperation and Development (2009e) Test guideline on subchronic inhalation toxicity: 90-day study, Paris
- OECD, Organisation for Economic Cooperation and Development (2009f) Test guideline on chronic toxicity studies, Paris
- OECD, Organisation for Economic Cooperation and Development (2010a) Guidelines for the testing of chemicals, section 4: health effects. Test no. 417: Toxicokinetics, Paris
- OECD, Organisation for Economic Cooperation and Development (2010b) Skin sensitization: local lymph node assay: DA, guidelines for testing of chemicals no. 442A, Paris
- OECD, Organisation for Economic Cooperation and Development (2010c) Skin sensitization: local lymph node assay: BrdU-ELISA, guidelines for testing of chemicals no. 442B, Paris
- OECD, Organisation for Economic Cooperation and Development (2010d) <http://www.oecd.org/dataoecd/33/37/37849783.pdf>. Accessed 30 June 2010
- OECD, Organisation for Economic Cooperation and Development (2010e) OECD guideline for the testing of chemicals draft proposal for a new guideline 4XX. The H295R steroidogenesis assay, Paris
- OECD, Organisation for Economic Co-operation and Development (2004c) Guidance document 28 for the conduct of skin absorption studies, Paris, March, 1–31
- Ohmori K, Umeda M, Tanaka N, Takagi H, Yoshimura I, Sasaki K, Asada S, Sakai A, Araki H, Asakura M, Baba H, Fushiwaki Y, Hamada S, Kitou N, Nakamura T, Nakamura Y, Oishi H, Sasaki S, Shimada S, Tsuchiya T, Uno Y, Washizuka M, Yajima S, Yamamoto Y, Yamamura E, Yatsushiro T (2005) Non-genotoxic carcinogen study group in the environmental mutagen society of Japan. An inter-laboratory collaborative study by the non-genotoxic carcinogen study group in Japan, on a cell transformation assay for tumour promoters using Bhas 42 cells. *Altern Lab Anim* 33:619–639
- Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Bracken W, Dorato M, Van Deun K, Sith P, Berger B, Heller A (2000) Concordance of the toxicity of pharmaceuticals in humans and animals. *Regul Toxicol Pharmacol* 32(1):56–67
- Oomen AG, Tolls J, Sips AJ, Groten JP (2003) In vitro intestinal lead uptake and transport in relation to speciation. *Arch Environ Contam Toxicol* 44(1):116–124
- Oosterhuis B (2010) Trends in microdosing and other exploratory human pharmacokinetic studies for early drug development. *Bioanalysis* 2(3):377–379
- Orman MA, Arai K, Yarmush ML, Androulakis IP, Berthiaume F, Ierapetritou MG (2010) Metabolic flux determination in perfused livers by mass balance analysis: effect of fasting. *Biotechnol Bioeng* 107(5):825–835
- Osman AM, van Dartel DA, Zwart E, Blokland M, Pennings JL, Piersma AH (2010) Proteome profiling of mouse embryonic stem cells to define markers for cell differentiation and embryotoxicity. *Reprod Toxicol* 30:322–332
- Ott H, Bergström MA, Heise R, Skazik C, Zwadlo-Klarwasser G, Merk HF, Baron JM, Karlberg AT (2009) Cutaneous metabolic activation of carboxime, a self-activating, skin-sensitizing pro-hapten. *Chem Res Toxicol* 22(2):399–405
- Ouwehand K, Santegoets SJ, Bruynzeel DP, Scheper RJ, de Gruijl TD, Gibbs S (2008) CXCL12 is essential for migration of activated Langerhans cells from epidermis to dermis. *Eur J Immunol* 38(11):3050–3059
- Ouwehand K, Spiekstra SW, Reinders J, Scheper RJ, de Gruijl TD, Gibbs S (2010) Comparison of a novel CXCL12/CCL5 dependent migration assay with CXCL8 secretion and CD86 expression for distinguishing sensitizers from non-sensitizers using MUTZ-3 Langerhans cells. *Toxicol In Vitro* 24(2):578–585
- Pant K, Sly JE, Bruce SW, Leung C, San RH (2008) Syrian hamster embryo (SHE) cell transformation assay with conditioned media (without X-ray irradiated feeder layer) using 2,4-diaminotoluene, 2,6-diaminotoluene and chloral hydrate. *Mutat Res* 654:108–113
- Paquette JA, Kumpf SW, Streck RD, Thomson JJ, Chapin RE, Stedman DB (2008) Assessment of the embryonic stem cell test and application and use in the pharmaceutical industry. *Birth Defects Res B Dev Reprod Toxicol* 83:104–111
- Park KB, Dalton-Brown E, Hirst C, Williams DP (2006) Selection of new chemical entities with decreased potential for adverse drug reactions. *Eur J Pharmacol* 549:1–8
- Parry JM, Parry E, Phrakonkham P, Corvi R (2010) Analysis of published data for top concentration considerations in mammalian cell genotoxicity testing. *Mutagenesis* 25:531–538
- Patlewicz GY, Worth A (2008) Review of data sources, QSARs and integrated testing strategies for skin sensitisation. *EUR 23225 EN*
- Patlewicz G, Aptula AO, Uriarte E, Roberts DW, Kern PS, Gerberick GF, Kimber I, Dearman RJ, Ryan CA, Basketter DA (2007) An evaluation of selected global (Q)SARs/expert systems for the prediction of skin sensitisation potential. *SAR QSAR Environ Res* 18(5–6):515–541
- Patlewicz G, Jeliaskova N, Gallegos Saliner A, Worth AP (2008) Toxmatch—a new software tool to aid in the development and evaluation of chemically similar groups. *SAR QSAR Environ Res* 19:397–412
- Pauwels M, Rogiers V (2009) The safety evaluation of cosmetic ingredients at the EU level and the use of 3R-alternatives: an update. Poster presentation in the “7th world congress on alternatives & animal use in the life sciences”, Rome, Italy. *ALTEX* 26(special issue):193
- Pauwels M, Rogiers V (2010) Human health safety evaluation of cosmetics in the EU: a legally imposed challenge to science. *Toxicol Appl Pharmacol* 243:260–274
- Pauwels M, Dejaegher B, Vander Heyden Y, Rogiers V (2009) Critical analysis of the SCCNFP/SCCP safety assessment of cosmetic ingredients (2000–2006). *Food Chem Toxicol* 47(4):898–905
- 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, 33 pp. http://www.ecb.jrc.ec.europa.eu/documents/REACH/PUBLICATIONS/REACH_testing_needs_final.pdf, JRC Report EUR 20863. Accessed 11 July 2010
- Peer Consultation on Health Canada Draft Weight of Evidence Framework for Genotoxic Carcinogenicity (2005) <http://www.tera.org/peer/comhaz/HC%20QSAR%20Peer%20Consultation%20Overview.pdf>
- Pelkonen O, Turpeinen M (2007) In vitro-in vivo extrapolation of hepatic clearance: biological tools, scaling factors, model assumptions and correct concentrations. *Xenobiotica* 37:1066–1089
- Pelkonen O, Turpeinen M, Uusitalo J, Rautio A, Raunio H (2005) Prediction of drug metabolism and interactions on the basis of in vitro investigations. *Basic Clin Pharmacol Toxicol* 96(3):167–175
- Pelkonen O, Kapitulnik J, Gundert-Remy U, Boobis AR, Stockis A (2008a) Local kinetics and dynamics of xenobiotics. *Crit Rev Toxicol* 38:697–720
- Pelkonen O, Tolonen A, Turpeinen M, Uusitalo J (2008b) In vitro metabolism in preclinical drug development. In: Gad SC (ed) *Preclinical development handbook: ADME and biopharmaceutical properties*, chapter 21. Wiley, New York, pp 743–774
- Pelkonen O, Turpeinen M, Hakkola J, Honkakoski P, Hukkanen J, Raunio H (2008c) Inhibition and induction of human cytochrome P450 enzymes—current status. *Arch Toxicol* 82:667–715
- Pelkonen O, Tolonen A, Korjamo T, Turpeinen M, Raunio H (2009a) From known knowns to known unknowns: predicting in vivo drug metabolites. *Bioanalysis* 1:393–414
- Pelkonen O, Tolonen A, Rousu T, Tursas L, Turpeinen M, Hokkanen J, Uusitalo J, Bouvier d'Yvoire M, Coecke S (2009b) Comparison of metabolic stability and metabolite identification of 55 ECVAM/ICCVAM validation compounds between human and rat liver homogenates and microsomes—a preliminary analysis. *ALTEX* 26(3):214–22
- Pendlington RU, Minter HJ, Stupart L, MacKay C, Roper CS, Sanders DJ, Pease CK (2008) Development of a modified in vitro skin absorption method to study the epidermal/dermal disposition of a contact allergen in human skin. *Cutan Ocul Toxicol* 27(4):283–294
- Péry AR, Brochot C, Desmots S, Boize M, Sparfel L, Fardel O (2010) Predicting in vivo gene expression in macrophages after exposure to benzo(a)pyrene based on in vitro assays and toxicokinetic/toxicodynamic models. *Toxicol Lett* 201(1):8–14
- Pessina A, Albella B, Bayo M, Bueren J, Brantom P, Casati S, Croera C, Gagliardi G, Foti P, Parchment R, Parent-Massin D, Schoeters G, Sibiril Y, Van Den Heuvel R, Gribaldo L (2003) Application of the CFU-GM assay to predict acute drug-induced neutropenia: an international blind trial to validate a prediction model for the maximum tolerated dose (MTD) of myelosuppressive xenobiotics. *Toxicol Sci* 75(2):355–367
- Pessina A, Malerba I, Gribaldo L (2005) Hematotoxicity testing by cell clonogenic assay in drug development and preclinical trials. *Curr Pharm Des* 11(8):1055–1065
- Pessina A, Bonomi A, Cavicchini L, Albella B, Cerrato L, Parent-Massin D, Sibiril Y, Parchment R, Behrsing H, Verderio P, Pizzamiglio S, Giangreco M, Baglio C, Coccè V, Sisto F, Gribaldo L (2010) Prevalidation of the rat CFU-GM assay for in vitro toxicology applications. *Altern Lab Anim* 38(2):105–117
- Peters AK, Wouwer GV, Weyn B, Verheyen GR, Vanparys P, Gompel JV (2008) Automated analysis of contractility in the embryonic stem cell test, a novel approach to assess embryotoxicity. *Toxicol In Vitro* 22:1948–1956
- Petrak J, Myslivcova D, Man P, Cmejla R, Cmejlova J, Vyoral D (2006) Proteomic analysis of iron overload in human hepatoma cells. *Am J Physiol Gastrointest Liver Physiol* 290(5):G1059–G1066
- Pfaffner W, Balls M, Clothier R, Coecke S, Dierickx P, Ekwall B, Hanley BA, Hartung T, Prieto P, Ryan MP, Schmuck G, Sladowski D, Vericat JA, Wendel A, Wolf A, Zimmer J (2001) Novel advanced in vitro methods for long-term toxicity testing: the report and recommendations of ECVAM workshop 45. European centre for the validation of alternative methods. *Altern Lab Anim* 29(4):393–426
- Pfuhler S, Kirkland D, Kasper P, Hayashi M, Vanparys P, Carmichael P, Dertinger S, Eastmond D, Elhajouji A, Krul C, Rothfuss A, Schoening G, Smith A, Speit G, Thomas C, van Benthem J, Corvi R (2009). Reduction of use of animals in regulatory genotoxicity testing: identification and implementation opportunities—report from an ECVAM workshop. *Mutat Res* 680:31–42
- Pfuhler S, Kirst A, Aardema M, Banduhn N, Goebel C, Araki D, Costabel-Farkas M, Dufour E, Fautz R, Harvey J, Hewitt NJ, Hibatallah J, Carmichael P, Macfarlane M, Reisinger K, Rowland J, Schelllauf F, Schepky A, Scheel J (2010) A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: genotoxicity. A COLIPA analysis. *Regul Toxicol Pharmacol* 57:315–324
- Piersma AH (2006) Alternative methods for developmental toxicity testing. *Basic Clin Pharmacol Toxicol* 98:427–431
- Podestà M, Piaggio G, Pitto A, Zocchi E, Soracco M, Frassoni F, Luchetti S, Painelli E, Bacigalupo A (2001) Modified in vitro conditions for cord blood-derived long-term culture initiating cells. *Exp Hematol* 29(3):309–314
- Poth A, Kunz S, Heppenheimer A (2007) Bhas cell transformation assay as a predictor of carcinogenicity. *ALTEX* 14(special issue):519–521
- Poulin P, Theil FP (2002) Prediction of pharmacokinetics prior to in vivo studies. 1. Mechanism-based prediction of volume of distribution. *J Pharmaceut Sci* 91:129–156
- Poulin P, Theil FP (2009) Development of a novel method for predicting human volume distribution at steady-state of basic drugs and comparative assessment with existing methods. *J Pharmaceut Sci* 98:4941–4961
- Prieto P (2002) Barriers, nephrotoxicology and chronic testing in vitro. *Altern Lab Anim* 30(suppl 2):101–105
- Prieto P, Baird AW, Blaauboer BJ, Ripoll JV, Corvi R, Dekant W, Dietl P, Gennari A, Gribaldo L, Griffin JL, Hartung T, Heindel JJ, Hoet P, Jennings P, Marocchio L, Noraberg J, Pazos P, Westmoreland C, Wolf A, Wright J, Pfaffner W (2006) The assessment of repeated dose toxicity in vitro: a proposed approach. The report and recommendations of ECVAM workshop 56. *Altern Lab Anim* 34(3):315–341
- Prieto P, Hoffmann S, Tirelli V, Tancredi F, González I, Bermejo M, De Angelis I (2010) An exploratory study of two caco-2 cell models for oral absorption: a report on their within-laboratory and between-laboratory variability, and their predictive capacity. *ATLA* 38(5):367–386
- Pritchard JB, French JE, Davis BJ, Haseman JK (2003) The role of transgenic mouse models in carcinogen identification. *Environ Health Perspect* 111:444–454
- Purdy R (1996) A mechanism-mediated model for carcinogenicity: model content and prediction of the outcome of rodent carcinogenicity bioassays currently being conducted on 25 organic chemicals. *Environ Health Perspect* 104:1085–1094
- Python F, Goebel C, Aeby P (2007) Assessment of the U937 cell line for the detection of contact allergens. *Toxicol Appl Pharmacol* 220(2):113–124
- Redfern WS, Ewart L, Hammond TG, Bialecki R, Kinter L, Lindgren S, Pollard CE, Roberts R, Rolf MG, Valentin JP (2010) Impact and frequency of different toxicities throughout the pharmaceutical life cycle (abstract 1081). *Toxicologist* 114:231
- Renwick AG (2004) Toxicology databases and the concept of thresholds of toxicological concern as used by the JECFA for

- the safety evaluation of flavouring agents. *Toxicol Lett* 149:223–234
- Ridings JE, Barratt MD, Cary R, Earnshaw CG, Eggington CE, Ellis MK, Judson PN, Langowski JJ, Marchant CA, Payne MP, Watson WP, Yih TD (1996) Computer prediction of possible toxic action from chemical structure: an update on the DEREK system. *Toxicology* 106:267–279
- Rietschel RL, Fowler JF (2008) Pathogenesis of allergic contact hypersensitivity. In: Fisher's contact dermatitis, 6th edn. BC Decker Inc., Hamilton, pp 1–10
- Riley RJ, Kenna JG (2004) Cellular models for ADMET predictions and evaluation of drug-drug interactions. *Curr Opin Drug Discov Dev* 7(1):86–99
- Ringerike T, Ullerås E, Völker R, Verlaan B, Eikeset A, Trzaska D, Adamczewska V, Olszewski M, Walczak-Drzewiecka A, Arkusz J, van Loveren H, Nilsson G, Lovik M, Dastych J, Vandebriel RJ (2005) Detection of immunotoxicity using T-cell based cytokine reporter cell lines ('cell chip'). *Toxicology* 206(2):257–272
- Rivedal E, Sanner T (1982) Promotional effect of different phorbol esters on morphological transformation of hamster embryo cells. *Cancer Lett* 17:1–8
- Roberts DW, Aptula AO (2008) Determinants of skin sensitisation potential. *J Appl Toxicol* 28(3):377–387
- Roberts DW, Patlewicz GY (2008) Nonanimal alternatives for skin sensitization: letter to the editor. *Toxicol Sci* 106(2):572–574
- Roberts DW, Patlewicz GY (2010) Updating the skin sensitization in vitro data assessment paradigm in 2009—a chemistry and QSAR perspective. *J Appl Toxicol* 30(3):286–288
- Roberts DW, Williams DL (1982) The derivation of quantitative correlations between skin sensitisation and physio-chemical parameters for alkylating agents, and their application to experimental data for sulfones. *J Theor Biol* 99(4):807–825
- Roberts DW, Aptula AO, Patlewicz G (2007) Electrophilic chemistry related to skin sensitisation. Reaction mechanistic applicability domain classification for a published data set of 106 chemicals tested in the mouse local lymph node assay. *Chem Res Toxicol* 20(1):44–60
- Roberts DW, Schultz T, Wolf E, Aptula AO (2010) Experimental reactivity parameters for toxicity modeling: application to the acute aquatic toxicity of SN2 electrophiles to *Tetrahymena pyriformis*. *Chem Res Tox* 23(1):228–234
- Robinson JF, van Beelen VA, Verhoef A, Renkens MF, Luijten M, van Herwijnen MH, Westerman A, Pennings JL, Piersma AH (2010) Embryotoxicant-specific transcriptomic responses in rat postimplantation whole-embryo culture. *Toxicol Sci* 118:675–685
- Rohrbeck A, Salinas G, Maaser K, Linge J, Salovaara S, Corvi R, Borlak J (2010) Toxicogenomics applied to in vitro carcinogenicity testing with Balb/c 3T3 cells revealed a gene signature predictive of chemical carcinogens. *Toxicol Sci* 118:31–41
- Rostami-Hodjegan A, Tucker GT (2007) Simulation and prediction of in vivo drug metabolism in human populations from in vitro data. *Nat Rev Drug Discov* 6:140–148
- Rotroff DM, Wetmore BA, Dix DJ, Ferguson SS, Clewell HJ, Houck KA, LeCluyse LE, Andersen ME, Judson RS, Smith CM, Sochaski MA, Kavlock RJ, Boellmann F, Martin MT, Reif DM, Wambaugh JF, Thomas RS (2010) Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening. *Toxicol Sci* 117:348–358
- Rougier N, Redziniak G, Mouglin D, Schmitt D, Vincent C (2000) In vitro evaluation of the sensitisation potential of weak contact allergens using Langerhans-like dendritic cells and autologous T cells. *Toxicology* 145(1):73–82
- Rowland M, Balant L, Peck C (2004) Physiologically based pharmacokinetics in drug development and regulatory science: a workshop report (Georgetown University, Washington, DC, May 29–30, 2002). *AAPS Pharm Sci* 6:E6
- Roy A, Georgopoulos PG (1998) Reconstructing week-long exposures to volatile organic compounds using physiologically based pharmacokinetic models. *J Expo Anal Environ Epidemiol* 8:407–422
- Sadikovic B, Al-Romaih K, Squire JA, Zielenska M (2008) Cause and consequences of genetic and epigenetic alterations in human cancer. *Curr Genom* 9:394–408
- Safford RJ (2008) The dermal sensitisation threshold—a TTC approach for allergic contact dermatitis. *Regul Toxicol Pharmacol* 51(2):195–200
- Sakaguchi H, Ashikaga T, Miyazawa M, Yoshida Y, Ito Y, Yoneyama K, Hirota M, Itagaki H, Toyoda H, Suzuki H (2006) Development of an in vitro skin sensitisation test using human cell lines; human cell line activation test (h-CLAT). II. An inter laboratory study of the h-CLAT. *Toxicol In Vitro* 20(5):774–784
- Sakaguchi H, Ashikaga T, Miyazawa M, Kosaka N, Ito Y, Yoneyama K, Sono S, Itagaki H, Toyoda H, Suzuki H (2009) The relationship between CD86/CD54 expression and THP-1 cell viability in an in vitro skin sensitisation test—human cell line activation test (h-CLAT). *Cell Biol Toxicol* 2592:109–126
- Sanderson DM, Earnshaw CG (1991) Computer prediction of possible toxic action from chemical structure; the DEREK system. *Hum Exp Toxicol* 10:261–273
- Sanner T, Dybing E (2005) Comparison of carcinogenic and in vivo genotoxic potency estimates. *Basic Clin Pharmacol Toxicol* 96:131–139
- Sawada H, Takami K, Asahi S (2005) A toxicogenomic approach to drug-induced phospholipidosis: analysis of its induction mechanism and establishment of a novel in vitro screening system. *Toxicol Sci* 83(2):282–292
- SCCNFP, European Commission's Scientific Committee on Cosmetics and Non-Food Products (2000) Opinion concerning the predictive testing of potentially cutaneous sensitising cosmetic ingredients or mixtures of ingredients, 17 February 2000
- SCCNFP, European Commission's Scientific Committee on Cosmetics and Non-Food Products (2003) Basic criteria for the in vitro assessment of dermal absorption of cosmetic ingredients, SCCNFP/0750/03. SCCNFP, Brussels
- SCCP, European Commission's Scientific Committee on Consumer's Products (2008) Opinion on dermal sensitisation quantitative risk assessment (citral, farnesol and phenylacetaldehyde), 24 June 2008
- SCCP, Scientific Committee on Consumer Products (2006) The SCCP's note of guidance for the testing of cosmetic ingredients and their safety evaluation, 6th revision, 106
- SCCS, European Commission's Scientific Committee on Consumer Safety (2009) Memorandum on alternative test methods in human safety assessment of cosmetic ingredients in the European Union, 8 December 2009
- Schaafsma G, Kroese ED, Tielemans EL, Van de Sandt JJ, Van Leeuwen CJ (2009) REACH, non-testing approaches and the urgent need for a change in mind set. *Regul Toxicol Pharmacol* 53(1):70–80
- Schaefer WR, Fischer L, Deppert WR, Hanjalic-Beck A, Seebacher L, Weimer M, Zahradnik HP (2010) In vitro-Ishikawa cell test for assessing tissue-specific chemical effects on human endometrium. *Reprod Toxicol* 30(1):89–93
- Schenk B, Weimer M, Bremer S, van der Burg B, Cortvrintd R, Freyberger A, Lazzari G, Pellizzer C, Piersma A, Schafer WR, Seiler A, Witters H, Schwarz M (2010) The ReProTect feasibility study, a novel comprehensive in vitro approach to detect reproductive toxicants. *Reprod Toxicol* 30:200–218
- Schmelzer E, Mutig K, Schrade P, Bachmann S, Gerlach JC, Zeilinger K (2009) Effect of human patient plasma ex vivo treatment on gene expression and progenitor cell activation of

- primary human liver cells in multi-compartment 3D perfusion bioreactors for extra-corporeal liver support. *Biotechnol Bioeng* 103(4):817–827
- Schmidt CW (2009) TOX 21: new dimensions of toxicity testing. *Environ Health Perspect* 117(8):A348–A353
- Schmitt W (2008) General approach for the calculation of tissue to plasma partition coefficients. *Toxicol In Vitro* 22:457–467
- Schneider K, Akkan Z (2004) Quantitative relationship between the local lymph node assay and human skin sensitisation assays. *Regul Toxicol Pharmacol* 39(3):245–255
- Schreiner M, Peiser M, Briechle D, Stahlmann R, Zuberbier T, Wanner R (2007) A loose-fit coculture of activated keratinocytes and dendritic cell-related cells for prediction of sensitising potential. *Allergy* 62(12):1419–1428
- Schroder HJ (1995) Comparative aspects of placental exchange functions. *Eur J Obstet Gynecol Reprod Biol* 63:81–90
- Schultz TW, Yarbrough JW, Johnson EL (2005) Structures-activity relationships for reactivity of carbonyl-containing compounds with glutathione. *SAR QSAR Environ Res* 16(4):313–322
- Schultz TW, Rogers K, Aptula A (2009) Read-across to rank skin sensitisation potential: subcategories for the Michael acceptor domain. *Contact Derm* 60(1):21–31
- Selderslaghs IW, Van Rompay AR, De CW, Witters HE (2009) Development of a screening assay to identify teratogenic and embryotoxic chemicals using the zebrafish embryo. *Reprod Toxicol* 28:308–320
- Sens-it-iv (2010) Novel testing strategies for in vitro assessment of allergens. Available at <http://www.sens-it-iv.eu/>. Accessed 14 July 2010
- Serafimova R, Todorov M, Pavlov T, Kotov S, Jacob E, Aptula A, Mekenyan O (2007) Identification of the structural requirements for mutagenicity, by incorporating molecular flexibility and metabolic activation of chemicals. II. General Ames mutagenicity model. *Chem Res Toxicol* 20:662–676
- Serafimova R, Fuat Gatnik M, Worth A (2010) Review of QSAR models and software tools for predicting genotoxicity and carcinogenicity. *EUR 24427 EN*. http://www.ecb.jrc.ec.europa.eu/DOCUMENTS/QSAR/EUR_24427_EN.pdf
- Shanks N, Greek R, Greek J (2009) Are animal models predictive for humans? *Philos Ethics Humanit Med* 4:1–20
- Shen O, Du G, Sun H, Wu W, Jiang Y, Song L, Wang X (2009) Comparison of in vitro hormone activities of selected phthalates using reporter gene assays. *Toxicol Lett* 191:9–14
- Smialowicz RJ (1995) Immune function testing for the identification and characterisation of immunotoxicity in rodents. *Hum Exp Toxicol* 14(1):135–136
- Smith CK, Hotchkiss SA (2001) *Allergic contact dermatitis*. Taylor & Francis, London, pp 1–16
- Smithing MP, Darvas F (1992) Hazardexpert: an expert system for predicting chemical toxicity. In: Finlay JW, Robinson SF, Armstrong DJ (eds) *Food safety assessment*. American Chem Society, Washington, DC, pp 191–200
- Snykers S, Vanhaecke T, Papeleu P, Luttun A, Jiang Y, Vander Heyden Y, Verfaillie C, Rogiers V (2006) Sequential exposure to cytokines reflecting embryogenesis: the key for in vitro differentiation of adult bone marrow stem cells into functional hepatocyte-like cells. *Toxicol Sci* 94:330–341
- Snykers S, De Kock J, Vanhaecke T, Rogiers V (2007) Differentiation of neonatal rat epithelial cells from biliary origin into immature hepatic cells by sequential exposure to hepatogenic cytokines and growth factors reflecting liver development. *Toxicol In Vitro* 21(7):1325–1331
- Snykers S, De Kock J, Rogiers V, Vanhaecke T (2009) In vitro differentiation of embryonic and adult stem cells into hepatocytes: state of the art. *Stem Cells* 27:577–605
- Sohn MD, McKone TE, Blancato JN (2004) Reconstructing population exposures from dose biomarkers: inhalation of trichloroethylene (TCE) as a case study. *J Expo Anal Environ Epidemiol* 14:204–213
- Spanhaak S, Cook D, Barnes J, Reynolds J (2008) Species concordance for liver injury: from the safety intelligence program board. BioWisdom, Ltd, Cambridge. Available at http://www.biowisdom.com/files/SIP_Board_Species_Concordance.pdf Accessed 15 Jan 2011
- Speit G (2009) How to assess the mutagenic potential of cosmetic products without animal tests? *Mutat Res* 678:108–112
- Spiekstra SW, dos Santos GG, Scheper RJ, Gibbs S (2009) Potential method to determine irritant potency in vitro—comparison of two reconstructed epidermal culture models with different barrier competency. *Toxicol In Vitro* 23(2):349–355
- Sreejit P, Kumar S, Verma RS (2008) An improved protocol for primary culture of cardiomyocyte from neonatal mice. *In Vitro Cell Dev Biol Anim* 44(3–4):45–50
- Steiner G, Suter L, Boess F, Gasser R, de Vera MC, Albertini S, Ruepp S (2004) Discriminating different classes of toxicants by transcript profiling. *Environ Health Perspect* 112(12):1236–1248
- Stemmer K, Ellinger Ziegelbauer H, Ahr HJ, Dietrich DR (2007) Carcinogen-specific gene expression profiles in short-term treated Eker and wild-type rats indicative of pathways involved in renal tumorigenesis. *Cancer Res* 67:4052–4068
- Stummann TC, Bremer S (2008) The possible impact of human embryonic stem cells on safety pharmacological and toxicological assessments in drug discovery and drug development. *Curr Stem Cell Res Ther* 3:118–131
- Stummann TC, Beilmann M, Duker G, Dumotier B, Fredriksson JM, Jones RL, Hasiwa M, Kang YJ, Mandenius CF, Meyer T, Minotti G, Valentin YJ, Zünlker BJ, Bremer S (2009a) Report and recommendations of the workshop of the European centre for the validation of alternative methods for drug-induced cardiotoxicity. *Cardiovasc Toxicol* 9(3):107–125
- Stummann TC, Hareng L, Bremer S (2009b) Hazard assessment of methylmercury toxicity to neuronal induction in embryogenesis using human embryonic stem cells. *Toxicology* 257:117–126
- Sutherland HJ, Lansdorp PM, Henkelman DH, Eaves AC, Eaves CJ (1990) Functional characterization of individual human hematopoietic stem cells cultured at limiting dilution on supportive marrow stromal layers. *Proc Natl Acad Sci USA* 87(9):3584–3588
- Suuronen EJ, Nakamura M, Watsky MA, Stys PK, Muller LJ, Munger R, Shinozaki N, Griffith M (2004) Innervated human corneal equivalents as in vitro models for nerve-target cell interactions. *FASEB J* 18(1):170–172
- Suzuki M, Morihiko H, Hagino S, Itagaki H, Aiba S (2009) Evaluation of changes of cell-surface thiols as a new biomarker for in vitro sensitisation test. *Toxicol In Vitro* 23(4):687–696
- Taléns-Visconti R, Bonora A, Jover R, Mirabet V, Carbonell F, Castell JV, Gómez-Lechón MJ (2006) Hepatogenic differentiation of human mesenchymal stem cells from adipose tissue in comparison with bone marrow mesenchymal stem cells. *World J Gastroenterol* 12(36):5834–5845
- Tan YM, Liao KH, Clewell HJ (2006a) Reverse dosimetry: interpreting trihalomethanes biomonitoring data using physiologically based pharmacokinetic modeling. *J Expo Anal Environ Epidemiol* 17:591–603
- Tan YM, Liao KH, Conolly RB, Blount BC, Mason AM, Clewell HJ (2006b) Use of a physiologically based pharmacokinetic model to identify exposures consistent with human biomonitoring data for chloroform. *J Toxicol Environ Health A* 69:1727–1756
- Tang W, Lu AYH (2010) Metabolic bioactivation and drug-related adverse effects: current status and future directions from a pharmaceutical research perspective. *Drug Metab Rev* 42:225–249

- Tatosian DA, Shuler ML (2009) A novel system for evaluation of drug mixtures for potential efficacy in treating multidrug resistant cancers. *Biotechnol Bioeng* 103:187–198
- Tennant RW, French JE, Spalding JW (1995) Identifying chemical carcinogens and assessing potential risk in short-term bioassays using transgenic mouse models. *Environ Health Perspect* 103:942–950
- Tennant RW, Stasiewicz S, Mennear J, French JE, Spalding JW (1999) Genetically altered mouse models for identifying carcinogens. *IARC Sci Publ* 146:123–150
- Testa B, Balmat AL, Long A (2004) Predicting drug metabolism: concepts and challenges. *Pure Appl Chem* 76(5):907–914
- Theunissen PT, Schulpen SH, van Dartel DA, Hermsen SA, van Schooten FJ, Piersma AH (2010) An abbreviated protocol for multilineage neural differentiation of murine embryonic stem cells and its perturbation by methyl mercury. *Reprod Toxicol* 29:383–392
- Thierbach R, Steinberg P (2009) Automated soft agar assay for the high-throughput screening of anticancer compounds. *Anal Biochem* 387:318–320
- Tilaoui L, Schilter B, Tran LA, Mazzatorta P, Grigorov M (2007) Integrated computational methods for prediction of the lowest observable adverse effect level of food-borne molecules. *QSAR Comb Sci* 26(1):102–108
- Tolonen A, Turpeinen M, Pelkonen O (2009) Liquid chromatography-mass spectrometry in in vitro drug metabolite screening. *Drug Discov Today (DDT)* 14:120–133
- Toropov AA, Toropova AP, Benfenati E (2009) Additive SMILES-based carcinogenicity models: probabilistic principles in the search for robust predictions. *Int J Mol Sci* 10:3106–3127
- ToxCast (2010) <http://www.epa.gov/nccst/toxcast/>
- Tsujimura K, Asamoto M, Suzuki S, Hokaiwado N, Ogawa K, Shirai T (2006) Prediction of carcinogenic potential by a toxicogenomic approach using rat hepatoma cells. *Cancer Sci* 97:1002–1010
- Turpeinen M, Ghiciuc C, Orpitoui M, Kankainen L, Pelkonen O, Pasanen M (2007) Predictive value of animal models for human cytochrome P450 (CYP)-mediated enzymes: a comparative study in vitro. *Xenobiotica* 37:1367–1377
- Uchino T, Taketzava T, Ikarashi Y (2009) Reconstruction of three-dimensional human skin model composed of dendritic cells, keratinocytes and fibroblasts utilizing a handy scaffold of collagen vitrigel membrane. *Toxicol In Vitro* 23(2):333–337
- Uehara T, Hirode M, Ono A, Kiyosawa N, Omura K, Shimizu T, Mizukawa Y, Miyagishima T, Nagao T, Urushidani T (2008) A toxicogenomics approach for early assessment of potential non-genotoxic hepatocarcinogenicity of chemicals in rats. *Toxicology* 250:15–26
- Uibel F, Muhleisen A, Kohle C, Weimer M, Stummann TC, Bremer S, Schwarz M (2009) ReProGlo: a new stem cell-based reporter assay aimed to predict embryotoxic potential of drugs and chemicals. *Reprod Toxicol* 30(1):103–112
- Urani C, Stefanini FM, Bussinelli L, Melchiorretto P, Crosta GF (2009) Image analysis and automatic classification of transformed foci. *J Microsc* 234:269–279
- US FDA (1995) Department of health and human services. Food additives; threshold of regulation for substances used in food-contact articles 21 CFR parts 5, 25, 170, 171, and 174
- US EPA, US Environmental Protection Agency (2004) In vitro dermal absorption rate testing of certain chemicals of interest to the occupational safety and health administration; final rule. *Fed Register* 69(80):22402–22441
- US EPA, US Environmental Protection Agency (2007a) Integrated summary report for the validation of an androgen receptor binding assay as a potential screen in the endocrine disruptor screening program. http://www.epa.gov/endo/pubs/ar_binding_isr.pdf. Accessed on 6 July 2010
- US EPA, US Environmental Protection Agency (2007b) Integrated summary report on aromatase. http://www.epa.gov/endo/pubs/aromatase_isr.pdf. Accessed on 6 July 2010
- US EPA, US Environmental Protection Agency (2009) Integrated summary report for validation of an estrogen receptor binding assay using rat uterine cytosol as source of receptor as a potential screen in the endocrine disruptor screening program tier 1 battery. http://www.epa.gov/endo/pubs/assayvalidation/er_ruc_isr.pdf. Accessed on 6 July 2010
- Vähäkangas K, Myllynen P (2006) Experimental methods to study human transplacental exposure to genotoxic agents. *Mutat Res* 608:129–135
- Van Calsteren V (2010) Analyse van de herhaalde toedieningstoxiciteitstesten met cosmetische ingrediënten op Europees niveau. Master thesis in pharmaceutical sciences. Dissertation, Vrije Universiteit Brussel
- Van Dartel DA, Pennings JL, Hendriksen PJ, van Schooten FJ, Piersma AH (2009) Early gene expression changes during embryonic stem cell differentiation into cardiomyocytes and their modulation by monobutyl phthalate. *Reprod Toxicol* 27:93–102
- Van Dartel DA, Pennings JL, van Schooten FJ, Piersma AH (2010) Transcriptomics-based identification of developmental toxicants through their interference with cardiomyocyte differentiation of embryonic stem cells. *Toxicol Appl Pharmacol* 243:420–428
- van Delft JH, van Agen E, van Breda SG, Herwijnen MH, Staal YC, Kleinjans JC (2005) Comparison of supervised clustering methods to discriminate genotoxic from non-genotoxic carcinogens by gene expression profiling. *Mutat Res* 575(1–2):17–33
- Van der Burg B, Winter R, Man HY, Vangenechten C, Berckmans P, Weimer M, Witters H, van der Linden S (2010a) Optimization and prevalidation of the in vitro AR CALUX method to test androgenic and antiandrogenic activity of compounds. *Reprod Toxicol* 30(1):18–24
- Van der Burg B, Winter R, Weimer M, Berckmans P, Suzuki G, Gijsbers L, Jonas A, van der Linden S, Witters H, Aarts J, Legler J, Kopp-Schneider A, Bremer S (2010b) Optimization and prevalidation of the in vitro ERalpha CALUX method to test estrogenic and antiestrogenic activity of compounds. *Reprod Toxicol* 30(1):73–80
- Van Leeuwen K, Schultz TW, Henry T, Diderich B, Veith GD (2009) Using chemical categories to fill data gaps in hazard assessment. *SAR QSAR Environ Res* 20:207–220
- van Loveren H, Cockshott A, Gebel T, Gundert-Remy U, de Jong WH, Matheson J, McGarry H, Musset L, Selgrade MK, Vickers C (2008) Skin sensitisation in chemical risk assessment: report of a WHO/IPCS international workshop focusing on dose–response assessment. *Regul Toxicol Pharmacol* 50(2):155–199
- Van Merris V, Van Wemmel K, Cortvrindt R (2007) In vitro effects of dexamethasone on mouse ovarian function and pre-implantation embryo development. *Reprod Toxicol* 23:32–41
- Van Och FM, Van Loveren H, Van Wolfswinkel JC, Machielsens AJ, Vandebriel RJ (2005) Assessment of potency of allergenic activity of low molecular weight compounds based on IL-1alpha and IL-18 production by a murine and human keratinocyte cell line. *Toxicology* 210(2–3):95–109
- Vandebriel RJ, van Loveren H (2010) Non-animal sensitisation testing: state-of-the-art. *Crit Rev Toxicol* 40(5):389–404
- Vanparys PH, Corvi R, Aardema M, Gribaldo L, Hayashi M, Hoffmann S, Schechtman L (2010) ECVAM prevalidation of three cell transformation assays. *ALTEX* 27:267–270
- Vedani A, Smiesko M (2009) In silico toxicology in drug discovery—concepts based on three-dimensional models. *Altern Lab Anim* 37:477–496

- Venkatapathy R, Moudgal CJ, Bruce RM (2004) Assessment of the oral rat chronic lowest observed adverse effect level model in TOPKAT, a QSAR software package for toxicity prediction. *J Chem Inf Comput Sci* 44(5):1623–1629
- Vermeire T, van de Bovenkamp M, de Bruin YB, Delmaar C, van Engelen J, Escher S, Marquat H, Meijster T (2010) Exposure-based waiving under REACH. *Regul Toxicol Pharmacol* 58(3):408–420
- Viravaidya K, Shuler ML (2004) Incorporation of 3T3-L1 cells to mimic bioaccumulation in a microscale cell culture analog device for toxicity studies. *Biotechnol Prog* 20:590–597
- Viravaidya K, Sin A, Shuler ML (2004) Development of a microscale cell culture analog to probe naphtalene toxicity. *Biotechnol Prog* 20:316–323
- Viswanath G, Halder S, Divya G, Majumder CB, Roy P (2008) Detection of potential (anti)progestagenic endocrine disruptors using a recombinant human progesterone receptor binding and transactivation assay. *Mol Cell Endocrinol* 295:1–9
- Vocanson M, Cluzel-Tailhardat M, Poyet G, Valeyrie M, Chavagnac C, Levarlet B, Courtellemont P, Rozières A, Hennino A, Nicolas JF (2008) Depletion of human peripheral blood lymphocytes in CD25+ cells allows for the sensitive in vitro screening of contact allergens. *J Invest Dermatol* 128(8):2119–2122
- Vocanson M, Hennino A, Rozišres A, Poyet G, Nicolas JF (2009) Effector and regulatory mechanisms in allergic contact dermatitis. *Allergy* 64(12):1699–1714
- Wainford RD, Weaver RJ, Hawksworth GM (2009) The immediate early genes, c-fos, c-jun and AP-1, are early markers of platinum analogue toxicity in human proximal tubular cell primary cultures. *Toxicol In Vitro* 23(5):780–788
- Walsh MJ, Bruce SW, Pant K, Carmichael PL, Scott AD, Martin FL (2009) Discrimination of a transformation phenotype in Syrian golden hamster embryo (SHE) cells using ATR-FTIR spectroscopy. *Toxicology* 258:33–38
- Wang J, Hou T (2009) Recent advances on in silico ADME modeling. *Annu Rep Comput Chem* 5:101–127
- Wanner R, Sonnenburg A, Quatchadze M, Schreiner M, Peiser M, Zuberbier T, Stahlmann R (2010) Classification of sensitising and irritative potential in a combined in vitro assay. *Toxicol Appl Pharmacol* 245(2):211–218
- Waters NJ, Jones R, Williams G, Sohal B (2008) Validation of a rapid equilibrium dialysis approach for the measurement of plasma protein binding. *J Pharmaceut Sci* 97:4586–4595
- Waters MD, Jackson M, Lea I (2010) Characterizing and predicting carcinogenicity and mode of action using conventional and toxicogenomics methods. *Mutat Res* 705:184–200
- Webborn PJ, Parker AJ, Denton RL, Riley RJ (2007) In vitro-in vivo extrapolation of hepatic clearance involving active uptake: theoretical and experimental aspects. *Xenobiotica* 37:1090–1099
- Weed D (2005) Weight of evidence: a review of concept and methods. *Risk Anal* 25(6):1545–1557
- Weiland C, Ahr HJ, Vohr HW, Ellinger-Ziegelbauer H (2007) Characterization of primary rat proximal tubular cells by gene expression analysis. *Toxicol In Vitro* 21(3):466–491
- West PR, Weir AM, Smith AM, Donley EL, Cezar GG (2010) Predicting human developmental toxicity of pharmaceuticals using human embryonic stem cells and metabolomics. *Toxicol Appl Pharmacol* 247:18–27
- Westmoreland C, Carmichael P, Dent M, Fentem J, MacKay C, Maxwell G, Pease C, Reynolds F (2010) Assuring safety without animal testing: unilever's ongoing research programme to deliver novel ways to assure consumer safety. *Altex 27 special issue from the world congress on alternatives to animals, Rome*, pp 61–65
- WHO, World Health Organisation (2008) World health statistics. <http://www.who.int/whosis/whostat/2008/en/print.html>. Accessed 15 Jan 2011
- Widschwendter M, Jones PA (2002) DNA methylation and breast carcinogenesis. *Oncogene* 21:5462–5482
- Wieser M, Stadler G, Jennings P, Streubel B, Pfaller W, Ambros P, Riedl C, Katinger H, Grillari J, Grillari-Voglauner R (2008) hTERT alone immortalizes epithelial cells of renal proximal tubules without changing their functional characteristics. *Am J Physiol Renal Physiol* 295(5):F1365–F1375
- Wilding I, Bell J (2005) Improved early clinical development through human microdosing studies. *Drug Discov Today* 10(13):890–894
- Willemsen P, Scippo ML, Kausel G, Figueroa J, Maghuin-Rogister G, Martial JA, Muller M (2004) Use of reporter cell lines for detection of endocrine-disrupter activity. *Anal Bioanal Chem* 378:655–663
- Williams DP (2006) Toxicophores: investigations in drug safety. *Toxicology* 226:1–11
- Wilmes A, Crean D, Aydin S, Pfaller W, Jennings P, Leonard MO (2010) Identification and dissection of the Nrf2 mediated oxidative stress pathway in human renal proximal tubule toxicity. *Toxicol In Vitro*. doi:10.1016/j.tiv.2010.12.009
- Winkler J, Sotiriadou I, Chen S, Hescheler J, Sachinidis A (2009) The potential of embryonic stem cells combined with -omics technologies as model systems for toxicology. *Curr Med Chem* 16:4814–4827
- Witters H, Freyberger A, Smits K, Vangenechten C, Lofink W, Weimer M, Bremer S, Ahr PH, Berckmans P (2010) The assessment of estrogenic or anti-estrogenic activity of chemicals by the human stably transfected estrogen sensitive MELN cell line: results of test performance and transferability. *Reprod Toxicol* 30(1):60–72
- Wolf T, Niehaus-Rolf C, Banduhn N, Eschrich D, Scheel J, Luepke NP (2008) The hen's egg test for micronucleus induction (HET-MN): novel analyses with a series of well-characterized substances support the further evaluation of the test system. *Mutat Res* 650:150–164
- Woo YT, Lai DY (2003) Mechanisms of action of chemical carcinogens, and their role in structure-activity relationships (SAR) analysis and risk assessment. In: Benigni R (ed) Quantitative structure-activity relationship (QSAR) models of mutagens and carcinogens. CRC Press, Boca Raton, pp 41–80
- Woo YT, Lai DY, Argus MF, Arcos JC (1995) Development of structure-activity relationship rules for predicting carcinogenic potential of chemicals. *Toxicol Lett* 79:219–228
- Wu S, Blackburn K, Amburgey J, Jaworska J, Federle T (2010) A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogs for SAR-based toxicological assessments. *Regul Toxicol Pharmacol* 56:67–81
- Xu X, Yang W, Li Y, Wang YH (2010) Discovery of estrogen receptor modulators: a review of virtual screening and SAR efforts. *Expert Opin Drug Discov* 5:21–31
- Yan L, Sheik-Bahaei S, Park S, Ropella GEP, Hunt CA (2008) Predictions of hepatic disposition properties using a mechanistically realistic, physiologically based model. *Drug Metab Dispos* 36:759–768
- Yang RSH, Thomas RS, Gustafson DL, Campaign J, Benjamin SA, Verhaar HJM, Mumtaz MM (1998) Approaches to developing alternative and predictive toxicology based on PBPK/PD and QSAR modeling. *Environ Health Perspect* 106:1385–1393
- Zachow R, Uzumcu M (2006) The methoxychlor metabolite, 2,2-bis-(p-hydroxyphenyl)-1,1,1-trichloroethane, inhibits steroidogenesis in rat ovarian granulosa cells in vitro. *Reprod Toxicol* 22:659–665
- Zaldívar JM, Mennecozzi M, Rodrigues R, Bouhifd M (2010) A biology-based dynamic approach for the modelling of toxicity in cell-based assays. Part I: fate modelling. EUR 24374 EN. European Commission, Joint Research Centre, Luxembourg

- Zanger UM, Turpeinen M, Klein K, Schwab M (2008) Functional Pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. *Anal Bioanal Chem* 392(6):1093–1108
- Zeilinger K, Schreiter T, Darnell M, Lübberstedt M, Müller-Vieira U, Dillner B, Urbaniak T, Knobloch D, Nüssler A, Gerlach JC, Andersson TB (2010) Human liver cell cultivation in a miniaturized 3D bioreactor system for in vitro studies on hepatic drug toxicity. *ALTEX* 27(suppl 2/10):148–149
- Zhang J, Musson DG (2006) Investigation of high-throughput ultrafiltration for the determination of an unbound compound in human plasma using liquid chromatography and tandem mass spectrometry with electrospray ionization. *J Chromatogr B* 843:47–56
- Zon LI, Peterson RT (2005) In vivo drug discovery in the zebrafish. *Nat Rev Drug Discov* 4:35–44
- Zuang V, Eskes C (2005) Alternative (non-animal) methods for cosmetics testing: current status and future prospects—a report prepared in the context of the 7th amendment of the cosmetic directive for establishing the timetable for phasing out animal testing—foreword. *Atla Altern Lab Anim* 33:1
- Zur Nieden NI, Ruf LJ, Kempka G, Hildebrand H, Ahr HJ (2001) Molecular markers in embryonic stem cells. *Toxicol In Vitro* 15:455–461
- Zur Nieden NI, Kempka G, Ahr HJ (2004) Molecular multiple endpoint embryonic stem cell test—a possible approach to test for the teratogenic potential of compounds. *Toxicol Appl Pharmacol* 194:257–269
- Zur Nieden NI, Davis LA, Rancourt DE (2010) Monolayer cultivation of osteoprogenitors shortens duration of the embryonic stem cell test while reliably predicting developmental osteotoxicity. *Toxicology* 277:66–73