Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

10-Aug-2012

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

PEER REVIEW REPORT RELATED TO THREE CELL TRANSFORMATION ASSAYS AND AGREEMENT OF THE WORKING GROUP OF THE NATIONAL COORDINATORS OF THE TEST GUIDELINES PROGRAMME ON THE FOLLOW-UP TO THE PEER REVIEW

Series on Testing and Assessment

No. 163

This document is only available in PDF format.

JT03325308

Complete document available on OLIS in its original format
This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

ENV/JM/MONO(2012)29
Unclassified
English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 163

PEER REVIEW REPORT RELATED TO THREE CELL TRANSFORMATION ASSAYS AND AGREEMENT OF THE WORKING GROUP OF THE NATIONAL COORDINATORS OF THE TEST GUIDELINES PROGRAMME ON THE FOLLOW-UP TO THE PEER REVIEW

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris 2012

Also published in the Series on Testing and Assessment:

No. 1, *Guidance Document for the Development of OECD Guidelines for Testing of Chemicals (1993; reformatted 1995, most recently revised 2009)*

No. 2, *Detailed Review Paper on Biodegradability Testing (1995)*

No. 3, *Guidance Document for Aquatic Effects Assessment (1995)*

No. 4, *Report of the OECD Workshop on Environmental Hazard/Risk Assessment (1995)*

No. 5, *Report of the SETAC/OECD Workshop on Avian Toxicity Testing (1996)*

No. 6, *Report of the Final Ring-test of the Daphnia magna Reproduction Test (1997)*

No. 7, *Guidance Document on Direct Phototransformation of Chemicals in Water (1997)*

No. 8, *Report of the OECD Workshop on Sharing Information about New Industrial Chemicals Assessment (1997)*

No. 9, *Guidance Document for the Conduct of Studies of Occupational Exposure to Pesticides during Agricultural Application (1997)*

No. 10, *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data (1998)*

No. 11, *Detailed Review Paper on Aquatic Testing Methods for Pesticides and industrial Chemicals (1998)*

No. 12, *Detailed Review Document on Classification Systems for Germ Cell Mutagenicity in OECD Member Countries (1998)*

No. 13, *Detailed Review Document on Classification Systems for Sensitising Substances in OECD Member Countries 1998)*

No. 14, *Detailed Review Document on Classification Systems for Eye Irritation/Corrosion in OECD Member Countries (1998)*

No. 15, *Detailed Review Document on Classification Systems for Reproductive Toxicity in OECD Member Countries (1998)*

No. 16, *Detailed Review Document on Classification Systems for Skin Irritation/Corrosion in OECD Member Countries (1998)*

No. 17, *Environmental Exposure Assessment Strategies for Existing Industrial Chemicals in OECD Member Countries (1999)*

No. 18, *Report of the OECD Workshop on Improving the Use of Monitoring Data in the Exposure Assessment of Industrial Chemicals (2000)*

No. 19, *Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (1999)*

No. 20, *Revised Draft Guidance Document for Neurotoxicity Testing (2004)*

No. 21, *Detailed Review Paper: Appraisal of Test Methods for Sex Hormone Disrupting Chemicals (2000)*

No. 22, *Guidance Document for the Performance of Out-door Monolith Lysimeter Studies (2000)*

No. 23, *Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures (2000)*

No. 24, *Guidance Document on Acute Oral Toxicity Testing (2001)*

No. 25, *Detailed Review Document on Hazard Classification Systems for Specifics Target Organ Systemic Toxicity Repeated Exposure in OECD Member Countries (2001)*

No. 26, *Revised Analysis of Responses Received from Member Countries to the Questionnaire on Regulatory Acute Toxicity Data Needs (2001)*

No 27, *Guidance Document on the Use of the Harmonised System for the Classification of Chemicals which are Hazardous for the Aquatic Environment (2001)*

No 28, *Guidance Document for the Conduct of Skin Absorption Studies (2004)*

No 29, *Guidance Document on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media (2001)*

No 30, *Detailed Review Document on Hazard Classification Systems for Mixtures (2001)*

No 31, *Detailed Review Paper on Non-Genotoxic Carcinogens Detection: The Performance of In-Vitro Cell Transformation Assays (2007)*

No. 32, *Guidance Notes for Analysis and Evaluation of Repeat-Dose Toxicity Studies (2000)*

No. 33, *Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures (2001)*

No. 34, *Guidance Document on the Development, Validation and Regulatory Acceptance of New and Updated Internationally Acceptable Test Methods in Hazard Assessment (2005)*

No. 35, *Guidance Notes for Analysis and Evaluation of Chronic Toxicity and Carcinogenicity Studies (2002)*

No. 36, *Report of the OECD/UNEP Workshop on the Use of Multimedia Models for Estimating Overall Environmental Persistence and Long Range Transport in the Context of PBTS/POPS Assessment (2002)*

No. 37, *Detailed Review Document on Classification Systems for Substances which Pose an Aspiration Hazard (2002)*

No. 38, *Detailed Background Review of the Uterotrophic Assay Summary of the Available Literature in Support of the Project of the OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA) to Standardise and Validate the Uterotrophic Assay (2003)*

No. 39, *Guidance Document on Acute Inhalation Toxicity Testing (2009)*

No. 40, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures which Cause Respiratory Tract Irritation and Corrosion (2003)*

No. 41, *Detailed Review Document on Classification in OECD Member Countries of Substances and Mixtures which in Contact with Water Release Toxic Gases (2003)*

No. 42, *Guidance Document on Reporting Summary Information on Environmental, Occupational and Consumer Exposure (2003)*

No. 43, *Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment (2008)*

No. 44, *Description of Selected Key Generic Terms Used in Chemical Hazard/Risk Assessment (2003)*

No. 45, *Guidance Document on the Use of Multimedia Models for Estimating Overall Environmental Persistence and Long-range Transport (2004)*

No. 46, *Detailed Review Paper on Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances (2004)*

No. 47, *Detailed Review Paper on Fish Screening Assays for the Detection of Endocrine Active Substances (2004)*

No. 48, *New Chemical Assessment Comparisons and Implications for Work Sharing (2004)*

No. 49, *Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs (2004)*

No. 50, *Report of the OECD/IPCS Workshop on Toxicogenomics (2005)*

No. 51, *Approaches to Exposure Assessment in OECD Member Countries: Report from the Policy Dialogue on Exposure Assessment in June 2005 (2006)*

No. 52, *Comparison of Emission Estimation Methods Used in Pollutant Release and Transfer Registers (PRTRs) and Emission Scenario Documents (ESDs): Case Study of Pulp and Paper and Textile Sectors (2006)*

No. 53, *Guidance Document on Simulated Freshwater Lentic Field Tests (Outdoor Microcosms and Mesocosms) (2006)*

No. 54, *Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application (2006)*

No. 55, *Detailed Review Paper on Aquatic Arthropods in Life Cycle Toxicity Tests with an Emphasis on Developmental, Reproductive and Endocrine Disruptive Effects (2006)*

No. 56, *Guidance Document on the Breakdown of Organic Matter in Litter Bags (2006)*

No. 57, *Detailed Review Paper on Thyroid Hormone Disruption Assays (2006)*

No. 58, *Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals (2006)*

No. 59, *Report of the Validation of the Updated Test Guideline 407: Repeat Dose 28-Day Oral Toxicity Study in Laboratory Rats (2006)*

No. 60, *Report of the Initial Work towards the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1A) (2006)*

No. 61, *Report of the Validation of the 21-Day Fish Screening Assay for the Detection of Endocrine Active Substances (Phase 1B) (2006)*

No. 62, *Final OECD Report of the Initial Work towards the Validation of the Rat Hershberger Assay: Phase-1, Androgenic Response to Testosterone Propionate, and Anti-Androgenic Effects of Flutamide (2006)*

No. 63, *Guidance Document on the Definition of Residue (2006)*

No. 64, *Guidance Document on Overview of Residue Chemistry Studies (2006)*

No. 65, *OECD Report of the Initial Work towards the Validation of the Rodent Uterotrophic Assay - Phase 1 (2006)*

No. 66, *OECD Report of the Validation of the Rodent Uterotrophic Bioassay: Phase 2. Testing of Potent and Weak Oestrogen Agonists by Multiple Laboratories (2006)*

No. 67, *Additional Data Supporting the Test Guideline on the Uterotrophic Bioassay in rodents (2007)*

No. 68, *Summary Report of the Uterotrophic Bioassay Peer Review Panel, including Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow up of this Report (2006)*

No. 69, *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models (2007)*

No. 70, *Report on the Preparation of GHS Implementation by the OECD Countries (2007)*

No. 71, *Guidance Document on the Uterotrophic Bioassay - Procedure to Test for Antioestrogenicity (2007)*

No. 72, *Guidance Document on Pesticide Residue Analytical Methods (2007)*

No. 73, *Report of the Validation of the Rat Hershberger Assay: Phase 3: Coded Testing of Androgen Agonists, Androgen Antagonists and Negative Reference Chemicals by Multiple Laboratories. Surgical Castrate Model Protocol (2007)*

No. 74, *Detailed Review Paper for Avian Two-generation Toxicity Testing (2007)*

No. 75, *Guidance Document on the Honey Bee (Apis Mellifera L.) Brood test Under Semi-field Conditions (2007)*

No. 76, *Final Report of the Validation of the Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances: Phase 1 - Optimisation of the Test Protocol (2007)*

No. 77, *Final Report of the Validation of the Amphibian Metamorphosis Assay: Phase 2 - Multi-chemical Interlaboratory Study (2007)*

No. 78, *Final Report of the Validation of the 21-day Fish Screening Assay for the Detection of Endocrine Active Substances, Phase 2: Testing Negative Substances (2007)*

No. 79, *Validation Report of the Full Life-cycle Test with the Harpacticoid Copepods Nitocra Spinipes and Amphiascus Tenuiremis and the Calanoid Copepod Acartia Tonsa - Phase 1 (2007)*

No. 80, *Guidance on Grouping of Chemicals (2007)*

No. 81, *Summary Report of the Validation Peer Review for the Updated Test Guideline 407, and Agreement of the Working Group of National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2007)*

No. 82, *Guidance Document on Amphibian Thyroid Histology (2007)*

No. 83, *Summary Report of the Peer Review Panel on the Stably Transfected Transcriptional Activation Assay for Detecting Estrogenic Activity of Chemicals, and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2007)*

No. 84, *Report on the Workshop on the Application of the GHS Classification Criteria to HPV Chemicals, 5-6 July, Bern Switzerland (2007)*

No. 85, *Report of the Validation Peer Review for the Hershberger Bioassay, and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2007)*

No. 86, *Report of the OECD Validation of the Rodent Hershberger Bioassay: Phase 2: Testing of Androgen Agonists, Androgen Antagonists and a 5 α-Reductase Inhibitor in Dose Response Studies by Multiple Laboratories (2008)*

No. 87, *Report of the Ring Test and Statistical Analysis of Performance of the Guidance on Transformation/Dissolution of Metals and Metal Compounds in Aqueous Media (Transformation/ Dissolution Protocol) (2008)*

No. 88, *Workshop on Integrated Approaches to Testing and Assessment (2008)*

No. 89, *Retrospective Performance Assessment of the Test Guideline 426 on Developmental Neurotoxicity (2008)*

No. 90, *Background Review Document on the Rodent Hershberger Bioassay (2008)*

No. 91, *Report of the Validation of the Amphibian Metamorphosis Assay (Phase 3) (2008)*

No. 92, *Report of the Validation Peer Review for the Amphibian Metamorphosis Assay and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2008)*

No. 93, *Report of the Validation of an Enhancement of OECD TG 211: Daphnia Magna Reproduction Test (2008)*

No. 94, *Report of the Validation Peer Review for the 21-Day Fish Endocrine Screening Assay and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-Up of this Report (2008)*

No. 95, *Detailed Review Paper on Fish Life-Cycle Tests (2008)*

No. 96, *Guidance Document on Magnitude of Pesticide Residues in Processed Commodities (2008)*

No. 97, *Detailed Review Paper on the use of Metabolising Systems for In Vitro Testing of Endocrine Disruptors (2008)*

No. 98, *Considerations Regarding Applicability of the Guidance on Transformation/Dissolution of Metals Compounds in Aqueous Media (Transformation/Dissolution Protocol) (2008)*

No. 99, *Comparison between OECD Test Guidelines and ISO Standards in the Areas of Ecotoxicology and Health Effects (2008)*

No. 100, *Report of the Second Survey on Available Omics Tools (2009)*

No. 101, *Report of the Workshop on Structural Alerts for the OECD (Q)SAR Application Toolbox, 15-16 May 2008, Utrecht, the Netherlands (2009)*

No. 102, *Guidance Document for using the OECD (Q)SAR Application Toolbox to Develop Chemical Categories According to the OECD Guidance on Grouping of Chemicals (2009)*

No. 103, *Detailed Review Paper on Transgenic Rodent Mutation Assays (2009)*

No. 104, *Performance Assessment: Comparsion of 403 and CxT Protocols via Simulation and for Selected Real Data Sets (2009)*

No. 105, *Report on Biostatistical Performance Assessment of the draft TG 436: Acute Toxic Class Testing Method for Acute Inhalation Toxicity (2009)*

No. 106, *Guidance Document for Histologic Evaluation of Endocrine and Reproductive Test in Rodents (2009)*

No. 107, *Preservative Treated Wood to the Environment for Wood Held in Storage after Treatment and for Wooden Commodities that are not Covered and are not in Contact with Ground. (2009)*

No. 108, *Report of the Validation of the Hershberger Bioassay (weanling model) (2009)*

No. 109, *Literature Review on the 21-Day Fish Assay and the Fish Short-Term Reproduction Assay (2009)*

No. 110, *Report of the Validation Peer Review for the Weanling Hershberger Bioassay and Agreement of the Working Group of National Coordinators of the Test Guidelines Programme on the Follow-up of this Report (2009)*

No. 111, *Report of the Expert Consultation to Evaluate an Estrogen Receptor Binding Affinity Model for Hazard Identification (2009)*

No. 112, *The 2007 OECD List of High Production Volume Chemicals (2009)*

No. 113, *Report of the Focus Session on Current and Forthcoming Approaches for Chemical Safety and Animal Welfare (2010)*

No. 114, *Performance Assessment of Different Cytotoxic and Cytostatic Measures for the In Vitro Micronucleus Test (MNVIT): Summary of results in the collaborative trial (2010)*

No. 115, *Guidance Document on the Weanling Hershberger Bioassay in Rats: A Short-term Screening Assay for (Anti) Androgenic Properties (2009)*

No. 116, *Guidance Document 116 on the Conduct and Design of Chronic Toxicity and Carcinogenicity Studies, Supporting Test Guidelines 451, 452 and 453 – 2nd Edition (2011)*

No. 117, *Guidance Document 117 on the Current Implementation of Internal Triggers in Test Guideline 443 for an Extended One Generation Reproductive Toxicity Study, in the United States and Canada (2011)*

No. 118, *Workshop Report on OECD Countries' Activities Regarding Testing, Assessment and Management of Endocrine Disrupters Part I and Part II (2010)*

No. 119, *Classification and Labelling of chemicals according to the UN Globally Harmonized System: Outcome of the Analysis of Classification of Selected Chemicals Listed in Annex III of the Rotterdam Convention (2010)*

No. 120, *Part 1: Report of the Expert Consultation on Scientific and Regulatory Evaluation of Organic Chemistry Mechanism-based Structural Alerts for the Identification of DNA Binding Chemicals (2010)*

No. 120, *Part 2: Report of the Expert Consultation on Scientific and Regulatory Evaluation of Organic Chemistry Mechanism-based Structural Alerts for the Identification of DNA Binding Chemicals (2010)*

No. 121, *Detailed Review Paper (DRP) on Molluscs Life-cycle Toxicity Testing (2010)*

No. 122, *Guidance Document on the Determination of the Toxicity of a Test Chemical to the Dung Beetle Aphodius Constans (2010)*

No. 123, *Guidance Document on the Diagnosis of Endocrine-related Histopathology in Fish Gonads (2010)*

No. 124, *Guidance for the Derivation of an Acute Reference Dose (2010)*

No. 125, *Guidance Document on Histopathology for Inhalation Toxicity Studies, Supporting TG 412 (Subacute Inhalation Toxicity: 28-Day) and TG 413 (Subchronic Inhalation Toxicity: 90-Day) (2010)*

No. 126, S*hort Guidance on the Threshold Approach for Acute Fish Toxicity (2010)*

No. 127, *Peer Review Report of the Validation of the 21-Day Androgenised Female Stickleback Screening Assay (2010)*

No. 128, *Validation Report of the 21-Day Androgenised Female Stickleback Screening Assay (2010)*

No. 129, *Guidance Document on Using Cytotoxicity Tests to Estimate Starting Doses for Acute Oral Systemic Toxicity Tests(2010)*

No. 131, *Report of the Test Method Validation of Avian Acute Oral Toxicity Test (OECD Test Guideline 223) (2010)*

No. 132, *Report of the Multi-Laboratory Validation of the H295R Steroidogenesis Assay to Identify Modulators (2010)*

No.133, *Peer Review Report for the H295R Cell-Based Assay for Steroidogenesis (2010)*

No.134, *Report of the Validation of a Soil Bioaccumulation Test with Terrestrial Oligochaetes by an International ring test (2010)*

No.135, *Detailed Review Paper on Environmental Endocrine Disruptor Screening: The use of Estrogen and Androgen Receptor Binding and Transactivation Assays in Fish (2010)*

No. 136, *Validation Report of the Chironomid Full Life-Cycle Toxicity Test (2010)*

No. 137, *Explanatory Background Document to the OECD Test Guideline on In Vitro Skin Irritation Testing (2010)*
No. 138, *Report of the Workshop on Using Mechanistic Information in Forming Chemical Categories (2011)*
No. 139, *Report of the Expert Consultation on Scientific and Regulatory Evaluation of Organic Chemistry Mechanism Based Structural Alerts for the Identification of Protein-binding Chemicals (2011)*
No. 140, *Report of the WHO/OECD/ILSI (Hesi) Workshop on Risk Assessment of Combined Exposures to Multiple Chemicals (2011)*
No. 141, *Report of the Phase 1 of the Validation of the Fish Sexual Development Test for the Detection of Endocrine Active Substances (2011)*
No. 142, *Report of the Phase 2 of the Validation of the Fish Sexual Development Test for the Detection of Endocrine Active Substances (2011)*
No. 143, *Peer Review Report for the Validation of the Fish Sexual Development Test and Agreement of the Working Group of National Co-ordinators of the Test Guideline Programme on the Follow-up of the Peer Review (2011)*
No. 144, *Validation Report for the Acute Chironomid Assay (2011)*
No. 145, *Transgenic Rodent Somatic and Germ Cell Gene Mutation Assay: Retrospective Performance Assessment (2011)*
No. 146, *Syrian Hamster Embryonic (SHE) Cell PH 6.7 Cell Transformation Assay Prevalidation Study Report (2012)*
No. 147, *Syrian Hamster Embryonic (SHE) Cell PH 7.0 Cell Transformation Assay Prevalidation Study Report (2012)*
No. 148, *Guidance Document on the Androgenised Female Stickleback Screen (2011)*
No. 149, *Validation Report of the Balb/c 3T3 Cell Transformation Assay (2012)*
No. 152, *Case Study: Assessment of an Extended Chemical Category, the Short-chain Methacrylates, Targeted on Bioaccumulation (2011)*
No. 153, *Guidance Document for the Derivation of an Acute Reference Concentration (Arfc) (2011)*
No. 154, *Validation Report: Part 1 – Validation of Efficacy Methods for Antimicrobials used on Hard Surfaces (2011)*
No. 154, *Validation Report: Part 2 – Validation of Efficacy Methods for Antimicrobials used on Hard Surfaces (2011)*
No. 155, *Peer Review for the Validation of the Modified Skin Irritation Test Method using LabyCyte EPI-MODEL24; Additional Studies; and Agreement of the Working Group of National Coordinators on the Follow-up to the Peer Review* (2011)
No. 156, *Guidance Notes on Dermal Absorption* (2011)
No. 157, *Validation Report Phase 1 for the Zebrafish Embryo Toxicity Test (2011)*

No. 158, *Report of Progress on the Interlaboratory Validation of the OECD Harpacticoid Copepod Development and Reproduction Test (2011)*
No. 159, *Validation Report for the Skin Irritation Test Method using Labcyte Epi-Model24 (2011)*
No. 160, *Guidance Document on the Bovine Corneal Opacity and Permeability (BCOP) and Isolated Chicken Eye (ICE) Test Methods: Collection of Tissues for Histological Evaluation and Collection of Data on Non-Severe Irritants (2011)*
No. 161, *Peer Review Report for the Validation of the Stably Transfected Transcriptional Activation Assay for the Detection of Androgenic and Anti-Androgenic Activity of Chemicals (2011)*
No. 162, *Guidance Document 116 on The Conduct And Design Of Chronic Toxicity And Carcinogenicity Studies, Supporting Test Guidelines 451, 452 And 453 – 2nd Edition (2011)*
No. 163, *Assays and Agreement of the Working Group of the National Coordinators of the Test Guidelines Programme on the Follow-up to the Peer Review (2012)*
No. 164, *Guidance Document on Crop Field Trials (2011)*
No. 165, *Report of an OECD Workshop on Metals Specificities in Environmental Risk Assessment (2011)*
No. 166, *SIDS Initial Assessment Profiles agreed in the course of the OECD HPV Chemicals Programme from 1993 to 2013 (2012)*
No. 167, *Crosswalk of Harmonized U.S. - Canada Industrial Function and Consumer and Commercial Product Categories with EU Chemical Product and Article Categories (2012)*
No. 168, *The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins Binding to Proteins (2012)*
No. 169, *Use of the AOP to Develop Chemical Categories and Integrated Assessment and Testing Approaches (2012)*

No. 170, *Draft Guidance Document for Demonstrating Efficacy of Pool and Spa Disinfectants and Field Testing (2012)*
No. 171, *Fish Toxicity Testing Framework (2012)*
No. 172, *Validity Report (Phase 2) for the Zebrafish Embryo Toxicity Test (2012)*
No. 173, *Performance Standards for Stably Transfected Transactivation in vitro Assays to Detect Estrogen Agonists for TG 455 Guidance Document on Standardised Test (2012)*
No. 174, *Performance Standards for the BG1Luc ER TA Transactivation Method to Detect Estrogen Receptor Antagonists (2012)*
No. 175, *Validation Report of a Ring Test for the OECD 305 Dietary Exposure Bioaccumulation Fish Test (Part 1) with Additional Report Including Comparative Analysis of Trout and Carp Results (Part II) (2012)*
No. 176, *Report on the Retrospective Analysis of 2-Generation Reprotoxicity Data (2012)*
No. 177, *Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption (2012)*

No. 178, *Detailed Review Paperstate of the Science on Novel In Vitro and In Vivo Screening and Testing Methods and Endpoints for Evaluating Endocrine Disruptors (2012)*
No. 179, *Draft Validation Report (Phase 2) for the Zebrafish Embryo Toxicity Test (2012)*

© OECD 2012
Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publications Service, RIGHTS@oecd.org. OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France

ABOUT THE OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 34 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in ten different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides and Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/ehs/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organisations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. UNDP is an observer. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

For this and many other Environment, Health and Safety publications, consult the OECD's World Wide Web site (www.oecd.org/ehs/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

FOREWORD

This document includes several documents related to *in vitro* Cell Transformation Assays using Syrian Hamster Embryo (SHE) cells (at pH 6.7 and pH 7.0) and the BALB/c 3T3 Mouse Fibroblast cell line:

- The ESAC[1] Working Group Peer Review Consensus Report on an ECVAM[2]-coordinated prevalidation study concerning three protocols of the Cell Transformation Assay (CTA) for *in vitro* carcinogenicity testing;

- The ESAC Opinion based on the above peer review;

- The EURL ECVAM[3] Recommendation on three Cell Transformation Assays using Syrian Hamster Embryo Cells (SHE) and the BALB/c 3T3 Mouse Fibroblast Cell Line for *In Vitro* Carcinogenicity Testing;

It also includes, in the front of the peer review report, the "Agreement of the Working Group of National Coordinators of the Test Guidelines Programme (WNT) on the follow-up to the peer review report on three Cell Transformation Assays".

The three prevalidation study reports for the CTAs using SHE cells (at pH 6.7 and pH 7.0) and the BALB/c 3T3 Cell Line for *In Vitro* Carcinogenicity Testing, submitted by ECVAM, have been published in the Series on Testing and Assessment as No. 146, 147, and 149. The above ESAC peer review report and the ESAC Opinion have been endorsed by the WNT at its April 2011 meeting.
The Secretariat prepared a draft WNT Agreement on the basis of the discussions and agreement reached at the expert group meeting that was held on 14-15 December 2011. As requested by the WNT at its 2011 meeting, the expert group reviewed all available information and discussed the applicability domain, and possible modes of action of the assays. It also discussed how the assays could be used and the need for specific targeted prospective testing. The draft WNT Agreement was slightly revised and approved by the WNT at its 2012 meeting; the Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology (hereafter Joint Meeting) agreed to its declassification on 7 August 2012.
This document is published under the responsibility of the Joint Meeting.

[1] ESAC : ECVAM Scientific Advisory Committee

[2] ECVAM : European Centre for the Validation of Alternative Methods

[3] EURL ECVAM: European Union Reference Laboratory for Alternatives to Animal Testing

Agreement of the Working Group of National Coordinators of the Test Guidelines Programme on the follow-up to the peer review report on three Cell Transformation Assays

In April 2011, the Working Group of National Coordinators of the Test Guidelines Programme (WNT) endorsed the three pre-validation reports for the SHE (pH 6.7 and 7.0) and the BALB/c3T3 cell transformation assays (CTAs), as well as the peer review report accompanied with the ESAC Opinion, and requested that experts meet to review all available information, clearly identify the applicability domain, discuss possible pathways and mechanisms of action of the assays, and discuss the need for specific targeted prospective testing. An expert group was held on 14-15 December 2011 at OECD. All information on CTAs (before and after the publication of the OECD Detailed Review Paper 31) was made available to the expert group.

The WNT, considering

- The CTAs' potential to improve chemical safety through the detection of both genotoxic and non genotoxic carcinogens;
- The CTAs' potential to reduce animal use in chemical testing;
- The extensive retrospective analysis included in the DRP 31 together with the ECVAM inter-laboratory study;
- The report and recommendation from the expert group held on 14-15 December 2011,

Agrees that

- a Test Guideline be developed for the SHE (Ph 6.7 and 7.0)
- a few more chemicals be tested with the BALB/c 3T3 to confirm the performance of the assay and the statistical approach used for data interpretation before a Test Guideline is developed.

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
European Centre for the Validation of Alternative Methods (ECVAM)

ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(ESAC)

ESAC Working Group Peer Review Consensus Report

on an ECVAM-coordinated prevalidation study concerning three protocols of the Cell Transformation Assay (CTA) for in vitro carcinogenicity testing

Title page information	
File name	TEMPLATE_ESAC-WG_REPORT.doc
Abbreviated title of ESAC request	ESAC peer review of and ESAC opinion on the ECVAM-led study of three cell transformation assays for carcinogenicity testing: 1) SHE pH 6.7 assay 2) SHE pH 7.0 assay 3) BALB/c 3T3 assay.
Relating to ESAC REQUEST Nr.	2010-02
Request discussed through	Plenary at ESAC33 at 2010/10/12
Report to be handed over to ESAC Chair and Secretariat by	25.01.2011
Version tracking	
File name	1102-18 ESAC WG CTA_approved.doc

TABLE OF CONTENTS

ESAC Working Group

This report was prepared by the "ESAC Working Group CTA" (ESAC WG), charged with conducting a detailed scientific peer review of the ECVAM prevalidation study concerning three protocols of the Cell Transformation Assay (CTA) for carcinogenicity testing.

The ESAC WG had been set up by the ESAC during its meeting on 12 October 2010. Basis for the scientific review was the ECVAM request to ESAC concerning the scientific review (ESAC request 2010-02, see Annex 5).

The ESAC WG conducted the peer review from 14 December 2010 to 25 January 2011. This report was endorsed by the ESAC WG on 16.2.2011 and represents the consensus view of the ESAC WG.

This ESAC WG peer review consensus report was endorsed by the ESAC on 18.2.2011.

The ESAC WG had the following members:

- Dr. Erwin ROGGEN (ESAC member, Chair of the ESAC Working Group)
- Dr. Rodger CURREN (ESAC member)
- Dr. David LOVELL (invited expert; EEP[1] member)
- Dr. Edgar RIVEDAL (invited expert; EEP member)
- Dr. Takeki TSUTSUI (invited expert following an ICATM proposal from JaCVAM; EEP member)

ESAC Secretariat:

- Dr. Claudius GRIESINGER (EC-ECVAM, ESAC Secretariat)
- Dr. Pascal PHRAKONKHAM (EC-ECVAM, specific support to ESAC Secretariat)

[1] EEP = ECVAM Expert Pool

NOTE ON THIS REPORTING TEMPLATE

The template follows the ECVAM modular approach and allows at the same time for the description of the analysis and conclusions concerning more specific questions. The template was approved by the ESAC through written procedure on 29 October 2010.

The template can be used for various types of validation studies (e.g. prospective full studies, retrospective studies, performance-based studies and prevalidation studies).

Depending on the study type and the objective of the study, not all sections may be applicable. However, for reasons of consistency and to clearly identify which information requirements have not been sufficiently addressed by a specific study, this template is uniformly used for the evaluation of validation studies.

- Explanatory notes to the paragraph titles (in green) have been added on 17 November 2010. These notes provide guidance on the type of information / analysis expected under each section. Depending on the purpose and scope of the study to be reviewed, some of the aspects mentioned in the explanatory notes may not be applicable or only be applicable to some extent. Moreover, the explanatory notes are not intended to represent an exhaustive list of possible issues to be addressed under the respective heading, but are thought to provide some guidance with respect to the considerations typically expected.
- For all of the template's numbered sections the summary view of ESAC WG is given in bold followed by more detailed comments ("general observations" and "specific observations").

ABBREVIATIONS USED IN THE DOCUMENT

- BLR Between-laboratory reproducibility
- CTA Cell Transformation Assay
- ECVAM European Centre for the Validation of Alternative Methods
- ESAC ECVAM Scientific Advisory Committee
- ESAC WG ESAC Working Group
- GCCP Good Cell Culture Practice
- GLP Good Laboratory Practice
- HCl Hydrochloric acid
- MTF Morphological Transformation Frequency
- OECD Organisation for Economic Cooperation and Development
- OECD DRP OECD Detailed Review Paper
- PC Positive Control
- SHE Syrian hamster embryo
- SOP Standard Operating Procedure (used here as equivalent to 'protocol')
- VC Vehicle Control
- VMT Validation Management Team
- WLR Within-laboratory reproducibility

Executive summary

Following a request from ECVAM to ESAC for peer review of and scientific advice on an ECVAM-coordinated prevalidation study concerning three protocol variants of the Cell Transformation Assay (CTA) used as in vitro predictors of carcinogenicity, an ESAC Working Group (ESAC WG) was set up by ESAC. The ESAC WG was charged with conducting a detailed scientific peer review of this study which had addressed protocol refinement of these three CTA protocols and assessed protocol transferability and reproducibility.

The ESAC WG met once in person at ECVAM in December 2010 and communicated further by email and teleconferences in December 2010 and January 2011. The ESAC WG reviewed the prevalidation study reports on the Syrian hamster embryo (SHE) pH6.7, SHE pH7.0 and BALB/c 3T3 CTA tests. The ESAC WG considered the scientific work presented was of good quality, despite some weaknesses in study design and execution.

The ESAC WG concluded that for both the SHE pH6.7 and SHE pH7.0 assays the results showed that a sufficiently standardised protocol is available which appears to be transferable, and that although data on within-laboratory variability was insufficient, there was reproducibility between laboratories on the basis of the six chemicals tested. The protocols were considered appropriate to serve as a basis for an OECD Guideline. However, future work using these protocols should aim at a complete characterisation of test method performance based on a larger set of chemicals.

Since a preliminary comparison made by the ESAC WG demonstrated appreciable similarity between the historical SHE protocols and those produced as a result of the prevalidation study (Annex 2), it is conceivable that historical data could be used to supplement or even substitute prospective testing data for the characterisation of test method performance. This characterisation should include information on reproducibility, predictive capacity, applicability and limitations of the SHE protocols.

The ESAC WG concluded that an improved BALB/c 3T3 CTA protocol had been developed. The WG considered that, although transferability had been achieved, within-laboratory reproducibility had not been sufficiently addressed and the between-laboratory reproducibility, while promising, had not been fully demonstrated. The protocol was considered suitable for further development towards the production of an OECD Guideline. However, the WG recommended that, prior to possible regulatory use there was a need to refine the acceptance and assessment criteria for the assay and to evaluate the test performance through dedicated test trials.

1. Data collection

1.1 Information / data sources used

NOTE: (Pre)validation studies typically make use of existing data, e.g. either as reference data (prospective studies) or as reference data and testing data as well (retrospective study). Have other data been used during the studies that were not generated during the study? If yes, for which purpose (e.g. reference data etc.)? What were that data sources?

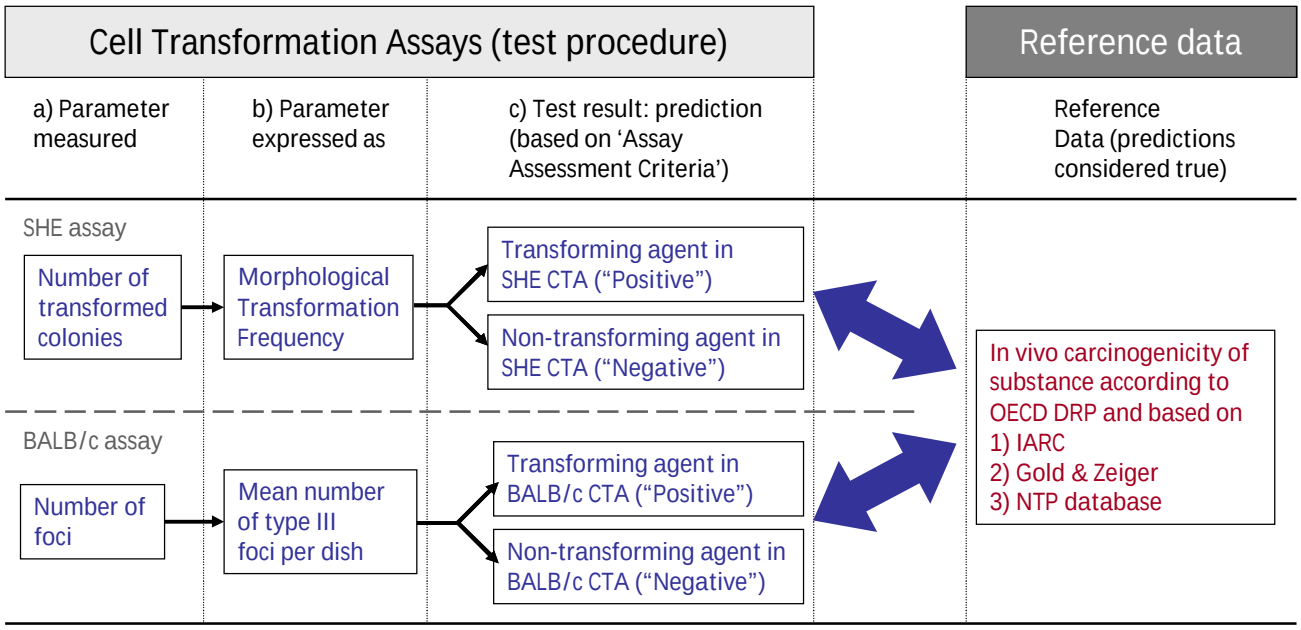
All reference data used in the prevalidation study are derived from the Detailed Review Paper on Cell Transformation Assays for Detection of Chemical Carcinogens published by the OECD in 2007 (hereunder abbreviated as "OECD DRP").

The current prevalidation study makes use, for each variant of the assay, of data from six of the chemicals reported in the OECD DRP as reference data to assess reproducibility as measured through the concordance of predictions within and between laboratories and in reference to in vivo carcinogenicity classifications as published in the OECD DRP, which in turn – in case of these chemicals – refers to classifications by the International Agency for Research on Cancer (IARC), Gold and Zeiger (1997) and the U.S. National Toxicology Program (NTP) database.

Briefly, the predictions generated by the Cell Transformation Assay (CTA) protocols in this study allow classification of the test chemical either as a "transforming agent" or as a "non-transforming agent" (for Syrian hamster embryo SHE, based on calculation of "Morphological Transformation Frequency", MTF; for BALB/c 3T3, based on measurement of the number of type III foci). Predictions (transforming/non-transforming agent) obtained in different laboratories were assessed for consistency (concordance) between the laboratories and were compared with in vivo carcinogenicity data as reported in the OECD DRP as the "reference standard"– see Figure 1.

Moreover, relevant ECVAM workshop reports and recent research papers related to (a) the scoring of observed effects, (b) the mechanistic understanding and (c) between-laboratory reproducibility are discussed in the prevalidation study reports (sections 1.4 to 1.7 pp.10-13 in all three reports), but have not been used as reference data.

Figure 1: Schematic depiction of the test procedure and the prediction generated by the SHE and BALB/c 3T3 CTAs and the reference data ("standard") used.



1.2 Search strategy

NOTE: How was the search for existing data planned, organised and executed? Has a search strategy been described and consistently applied?

The ESAC working group (ESAC WG) noted that there was apparently no detailed search strategy established for identifying suitable reference data. However, taking into account that this was a small-scale prevalidation study which did not attempt to define the predictive capacity or the applicability domain of the three CTAs studied, but focused on protocol refinement and reproducibility, this fact was not considered relevant in this context.

The information referred to in the CTA prevalidation study was extracted from the OECD DRP and, for purposes of discussing the background and status quo of the methods, from reports of ECVAM workshops and to a lesser extent from recent literature.

The OECD DRP was the sole source for reference data used for assessing within- and between-laboratory reproducibility (measured through concordance of predictions obtained in different laboratories in reference to published in vivo carcinogenicity predictions).

It was noted by the ESAC WG that the OECD DRP does not specify (1) what the search strategy for retrieving literature was, (2) whether there were selection criteria (e.g. based on study quality) to select or reject retrieved data and (3) whether studies that were originally retrieved had been indeed rejected because of quality issues (see point 2). This may affect the quality of the data and information represented in the OECD DRP and thus the quality of the reference data in the prevalidation study.

1.3 Selection criteria applied to the available information

NOTE: Have consistent evaluation/decision criteria been pre-defined and applied in order to select the data and has data selection been explained in a transparent manner?

The ESAC WG noted that there was apparently no detailed set of selection criteria established to reject/accept retrieved data. The OECD DRP was taken as reliable source although it is not clearly described in the OECD DRP how the quality of the data had been controlled.

The ESAC WG noted that for the purposes of the current prevalidation study, CTA test data from the OECD DRP and relating to six chemicals were selected on the basis of criteria established by the Validation Management Team (VMT) for selecting test chemicals (section 2.4.1 p.16 in all three reports). Once the chemicals had been identified, the relevant CTA reference data were extracted from the OECD DRP, apparently without further critical review of the data quality. Thus, there was no quality control of the reference data per se.

The OECD DRP does not mention whether the data reported were analysed for their quality and whether some of the data retrieved were rejected (i.e. on the basis of predefined minimum criteria for study quality). Thus, the issue of a lack of control on the reference data may have been propagated through to the prevalidation study.

2. Study objective and design

2.1 Clarity of the definition of the study objective

NOTE: Is the objective of the study clearly and comprehensibly defined?

The objective of the studies was considered clear and comprehensive: standardisation of CTA protocols and subsequent assessment of these protocols for reproducibility and transferability.

The objective was to establish a harmonized standard operation procedure (SOP) for each of the selected CTAs and to assess whether these method protocols are reproducible within laboratories, transferable and reproducible between laboratories.

It is stated in the study objective, that this information may facilitate the development of an OECD guideline based on these tests. The precise purpose of a possible future test guideline using CTAs is, however, not further elaborated on. In the opinion of the ESAC WG this may have affected the study design. For instance, considerations of regulatory needs, for instance the need to have additional information on predictive capacity, could have influenced chemical selection. Chemicals may have been selected based more on the amount of information available for their carcinogenic potential than on their potential to be good challenges for a reproducibility study.

2.2 Analysis of the scientific rationale provided

NOTE: Is the scientific rationale for the test method AND (consequently) for conducting the study clearly explained? How does the test method contribute scientifically to the scientific understanding / prediction of the specified health/environmental effect or aspects of it?

The intended scientific rationale was explained as far as our current understanding of the cellular mechanisms involved in carcinogenesis (primarily in rodent cells) allows. The reported prevalidation study does not contribute to this scientific understanding, but builds upon evidence (provided primarily by the OECD DRP) that genotoxic as well as non-genotoxic carcinogens induce cell transformation in SHE and BALB/c 3T3 cells while non-carcinogenic substances do not.

The prevalidation study reports discuss the scientific background in some detail (sections 1.1 to 1.4 pp.8-11 in all three reports). According to this brief overview, the CTAs could be used as a cell-based tool for carcinogenic hazard identification.

As outlined in the reports, the CTAs have been shown to involve a multistage cellular process that closely models some stages of in vivo carcinogenesis and have the potential to detect both genotoxic and non-genotoxic carcinogens. In addition, they are conducted to screen for potential carcinogenicity as well as to investigate mechanisms of carcinogenicity of industrial chemicals, agro-chemicals, cosmetics, pharmaceuticals, environmental hormones (xenobiotics), exhaust gas etc.

The suitability of the test for expanding the scientific understanding (i.e. studying in detail mechanisms of action) is discussed to some extent in the prevalidation study reports (section 1.4

pp.10-11 in all three reports). This is exemplified by mechanistic studies of di(2-ethylhexyl)phthalate genotoxicity in SHE pH6.7 and SHE pH7.0 CTAs (last paragraph of section 1.4, p.11 of both prevalidation study reports).

2.3 Analysis of the regulatory rationale provided

NOTE: Is a regulatory rationale specified, i.e. a specific application of the test method for purposes of generating data with respect to regulatory requirements as specified in legislation or internationally agreed guidelines etc.? If so, how does the study and its objective and design relate to this regulatory rationale?

The regulatory rationale remains somewhat open although it is acknowledged by the ESAC WG that even screening data and supportive data within a weight of evidence framework can be used for regulatory purposes and may thus constitute a "regulatory rationale". However, recommendations for a more precise definition of the regulatory usability of these tests should have been made in the reports since such use is mentioned as one of the motives for the study (see also section 15).

The current study was performed from 2005 to 2010 and planned on the background of past ECVAM activities towards the use of CTAs (Combes et al., 1997) and the OECD project of developing an OECD DRP which took place from 1997 to 2007. Both projects therefore overlapped to some extent and information on progress in both projects was mutually taken into consideration: while the ECVAM studies relied to a great extent on reference data compiled in the OECD DRP, the recommendations made in the OECD DRP took already into account possible results of the prevalidation study conducted by ECVAM at that time and aiming at the refinement of selected CTA protocols.

According to these recommendations of the OECD DRP, the reported studies (if successful) should facilitate the incorporation of the tests into OECD guidelines by demonstrating their reproducibility and transferability.

However, the specific purpose of the tests within the framework of, for example, an OECD test guideline was neither defined in the OECD DRP nor by the VMT when planning the current study.

2.4 Appropriateness of the study design

NOTE: This includes an analysis of the selection of test items, the number of test items, the number of laboratories involved in the study, retesting in case of unqualified tests and other technical aspects of the study.

Overall, the study design was considered appropriate for assessing the reproducibility and transferability of the standardised protocols, despite shortcomings relating to the design/planning of (1) the test item selection (even when considering that this is a prevalidation study, there are concerns regarding the number representativeness of test chemicals with regard to chemical class and mechanism of action), (2) the within-laboratory variability phase and (3) the transferability phase.

General observations:

(a) Test item selection: The selection criteria as defined and applied for choosing the six test items (chemical substances) may lead to a preponderance of strong positive/strong negative chemicals, which could give more reproducible results across different laboratories (section 2.4.1.1 p.16 in all three reports). In general, more than six substances should be used in a prevalidation study to assess reproducibility, although the number is sufficient to allow statistical evaluation of reproducibility. A more detailed discussion of the ESAC WG concerning number and representativeness of test items is provided in sections 5.1 and 5.2, respectively.

(b) Within-laboratory reproducibility: Only one chemical was used to assess the within-laboratory reproducibility. Moreover, this chemical was the positive control chemical, which increases the likelihood of obtaining reproducible results.

(c) Transfer and transferability: In agreement with the provisions for prevalidation (ECVAM 1995; OECD 2005), transfer of the protocol (formal prevalidation phase II) was ensured through appropriate measures (e.g. training, generation of photo catalogues for consistent scoring).

The reports furthermore indicate that the "transferability module" was assessed (e.g. p.9, 2nd paragraph in the SHE pH7.0 CTA report). This statement is however not substantiated by the study design and conduct: since all laboratories involved in the prevalidation studies had substantial experience with performance of the SHE and BALB/c 3T3 CTAs and since performance in such naive laboratories was not assessed through empirical testing of, at least a subset, of the test items, ease of transferability (which is part of the "transferability" assessment) was in fact not assessed.

Two issues should be acknowledged in this context:

(1) While testing the ease of transferability is not a specific requirement of prevalidation studies (ECVAM 1995; OECD Guidance Document Nr. 34[2]), it may be considered good practice in validation to include such a testing phase to give information about the possible

[2] The guidance document outlines that "Additional activities at the prevalidation stage may include an assessment of the transferability of the test protocol to laboratories inexperienced in the test or the necessary techniques and to resolve questions or inconsistencies arising in this phase" (Paragraph 75, OECD Guidance Document Nr. 34 on the Validation and International Acceptance of new or updated test methods for hazard assessment, OECD 2005)

problems that may arise in transferring a method and/or its robustness in view of possible minor (uncontrolled) alterations of test procedure. In particular, when considering that the test methods depend on visual scoring, data on ease of transferability would have been helpful.

(2) It should be noted however, that the VMT has obviously given this issue some thought and the ESAC WG acknowledges that the reports discuss, for example, to which extent within-laboratory variability data of the different laboratories (and all based on testing the positive control only) could be useful in making conclusions on ease of transferability. See for instance sections 2.4.2 p.17 and 2.6, 5th paragraph on pp.18-19 in the SHE pH7.0 CTA report, the latter reproduced below:

"Following the preliminary phase of optimisation of the protocols, the transferability and the within- laboratory reproducibility were assessed by evaluating results obtained for one non-coded test chemical and a coded one. These two chemicals were the same (Benzo(a)pyrene), allowing an analysis of the within-laboratory reproducibility as well as the transferability of the assay."

(d) The number of laboratories involved in each study was considered sufficient (n=3-4) (section 2.2 p.15 in all three reports).

(e) While it is recognized that using the same cell batch in all laboratories involved in a study is good practice in the context of validation (section 3.5.1 p.21 and p.22 in the SHE pH6.7 CTA and SHE pH7.0 CTA reports, respectively), this does not reflect the real testing situation once the tests are used in applied safety testing/research. It, therefore, does not provide information on the extent to which batch-to-batch variability may influence the reproducibility of these tests.

(f) The studies were performed by GLP-certified laboratories, but not under GLP (Good Laboratory Practice). The use of the expression "GLP-like" (section 2.3 p.15 in all three reports) can be criticised as potentially implying a quality-controlled study, while this can not be derived from the mere fact that some laboratories participating in the studies have GLP certification and work under GLP. The ESAC WG holds that it should have been possible to state in which ways the study conduct differed from full compliance with GLP.

Specific observations:

(a) SHE pH7.0 CTA report, p.22, section 3.5.4 Controls:

The reason why the concentration of benzo(a)pyrene was different between the University of Metz and the other laboratories should be specified.

(b) SHE pH7.0 CTA report, p.53, Figure 14:

The University of Metz obtained high MTFs with o-toluidine HCl at 20 •g/ml and 40 •g/ml. The VMT concluded that a statistical significant MTF at the lower doses could be related to the University of Metz using a different batch of cells and serum than the other laboratories. However, in order to confirm this, it may be necessary to re-check the colonies by investigators from other laboratories. It is possible that the University of Metz used criteria different from those used by other laboratories but this has not been specified.

(c) In the BALB/c 3T3 CTA report, p.26, section 3.5.3 Conclusion of the Validation Management Team on preliminary experiments, the VMT stated that “the maximum number of Type III foci in the entire set of vehicle control dishes should not exceed five”. The specific reason for requiring a maximum of “five” foci should be explained and justified.

2.5 Appropriateness of the statistical evaluation

NOTE: Are the statistical methods used for evaluating the study data appropriate. Is there a sufficient justification for the use of the methods chosen? Was the statistician independent from the test method submitter/developer?

The statistical evaluation of the test data generated during the study appears appropriate.

However, the methods of statistical analysis used in the test method procedures (SOP) and the assay assessment criteria need a critical revision as outlined below:

General observations:

(a) Plate variability is not taken into account by the Fisher exact test, i.e. Fisher exact test assumes the colony is the experimental unit (and not the plate).

(b) The use of the negative binomial analysis should be critically assessed and justified. Evidence that it gives comparable results to those using the Fisher exact test should be provided.

(c) Anomalous results in the tables need explanations.

(d) Individual plate data should be available.

Specific observations:

For all reports:

(a) Terminology issues: Care is needed as there are two different uses of the term transformation: statistical transformation of data and cell transformation by a series of stages (in this context assessed as morphological transformation).

(b) Statistical analysis seems to have been independent of the conduct of the study and has been carried out "blinded". However, there is some subjectivity or expert judgement applied to the final decision. The criteria for determining a positive result were not completely followed with the use of other information to overrule criteria that were not met. It was noted that a lack of statistical significance was in some cases overridden by a positive biological interpretation which suggests a low power for the studies and/or appreciable uncertainty over what constitutes a positive or negative result.

(c) Note that a statistical analysis is not "negative". A statistical test relates to acceptance or rejection of a null hypothesis. A "negative" result means that there is insufficient evidence to reject the Null Hypothesis. It is not that the Null Hypothesis is true.

For the SHE pH6.7 and SHE pH7.0 CTA reports:

(d) The statistical methods for the SHE pH6.7 and SHE pH7.0 CTAs are simple (section 3.5.6 p.23 and p.24 in the SHE pH6.7 CTA and SHE pH7.0 CTA reports, respectively). In spite of being the standard test in the context of the SHE CTA methods, there are some limitations to the use of the Fisher exact test and it may not be the most appropriate statistical method.

(e) The decision rule using the Fisher exact test is "any two comparisons significant" and a Cochran-Armitage trend test (if only one Fisher exact test is significant). The statistical criteria for a significant result is if P <0.05. No correction was made for multiple comparisons.

(f) The decision rule used could lead to some idiosyncratic "calls". It assumes, for instance, that the dose-response is linear (different from "general positive trend" (p.23 and p.24 in the SHE pH6.7 and pH7.0 CTA reports, respectively)). It is not exactly clear what the statistical properties of this decision rule are. As the decision rule is defined there is no specific dose relationship needed between the two significant dose comparisons required for a "positive" call.

(g) The Fisher (not Fischer! p.38 in the BALB/c 3T3 CTA report) exact test is vulnerable to false positives in that it expects each unit to be independent. In fact, it is not clear how many colonies are independent of one another. It is unclear how many plates are independent. "Rogue" plates could make an effect seemingly statistically significant.

(h) It is unclear whether zero results (e.g. table 9 p.36 for 2.5 µg/ml and table 30 p.57 for 55 and 110 µg/ml in the SHE pH7.0 CTA report) considered acceptable results.

(i) It is not clear that a "complete concordance" of results of statistical conclusions (section 6 p.34 and p.35 in the SHE pH6.7 and pH7.0 CTA reports, respectively) is a "data quality control check". Two copies of the same data may be identical but that does not mean that the data are of high quality.

<u>For the BALB/c 3T3 CTA report:</u>

(j) The acceptable number of vehicle control foci is <6. This should be justified in more detail.

(k) The recommended statistical methods suggested for the BALB/c 3T3 CTA (section 3.6.5.3 pp.29-30 in the BALB/c 3T3 CTA report) are complex and have not been fully investigated to allow use in practice.

(l) These statistical methods are somewhat specialized and not widely used. The exact properties of the methods are not defined and need some discussion. The methods are based upon either the Negative Binomial combined with a Williams-type test or the Nishiyama transformation. The only reference to this transformation seems to be a paper by Nishiyama (section 3.6.5.2, 3rd paragraph p.29 in the BALB/c 3T3 CTA report).

(m) For the within-laboratory reproducibility phase (section 4 p.32), the text is ambiguous because the statistical analysis is described as being initially by Fisher exact tests (section 4.3 p.37), which were superseded by the Negative Binomial combined with a Williams-type test after the statistician ad hoc expert meeting, and both approaches "gave comparable results" with 3-methylcholanthrene. This seems unlikely given the properties of the different tests (i.e. a series of pair-wise comparisons for Fisher tests vs. identification of a significant trend for the Negative Binomial combined with a Williams-type test), although the calls based upon the decision criteria set for each test were found to be the same. It may be that this statement relates, in fact, to giving comparable "calls" given the decision criteria used (table 9 p.37).

(n) The provision of the code for the "R programme" is good practice (annex 13.3.2 pp.101-104 in the BALB/c 3T3 CTA report).

(o) It is not clear how significant result can be obtained when values at dose level are the same as those with the vehicle control (e.g. in table 10 p.41, 1.2 vs. 1.2 foci per dish) or lower than the vehicle control (e.g. in table 14 p.47, 0.20 vs. 0.40 foci per dish). These are indicated and described as having "contributed to a significant result i.e. which were present in the

downward-protected Williams contrast that resulted in the lowest p-value (negative binomial analysis) ...". This may be a feature of the decision criteria used. This appears to relate to the criteria for the detection of a positive trend which can include results from dose levels which are as low or lower than the vehicle control values. Reporting positive effects partly based upon values at or below the vehicle control values can, however, be difficult to envisage and may be liable to misinterpretation.

3. Test definition (Module 1)

3.1 Quality and completeness of the overall test definition

NOTE: This included an analysis of the description of the test system, the protocol, test acceptance criteria etc.

Overall the tests were adequately defined considering the objective of this study.

Specific observations:

For the SHE pH6.7 and SHE pH7.0 CTA reports:

(a) The endpoint is the number of transformed colonies defined by cells with altered morphology and disorganized patterns of cell growth (a potential "subjectivity" issue).

(b) The test system is well described as far as the current mechanistic understanding allows. The reasons for changing the pH from 7.0 to 6.7 is adequately explained (specific for SHE pH6.7 CTA, section 3.4 p.21 in the SHE pH6.7 CTA report).

(c) The protocol is detailed and clear (section 3.5 pp.21-24 and pp.22-25 in the SHE pH6.7 CTA and SHE pH7.0 CTA reports, respectively), and addressing the important issues (e.g. medium, serum) individually. The controls were specified.

(d) Acceptance criteria are listed and are clear (section 3.5.7 p.23 and pp.24-25 in the SHE pH6.7 CTA and SHE pH7.0 CTA reports, respectively).

(e) Section 9 Recommendation, 2nd paragraph p.61 and p.64, in the SHE pH6.7 CTA and SHE pH7.0 CTA reports, respectively: “since they only differ by the pH used to culture…” should be “since they only differ by the pH and medium used to culture…”.

For the BALB/c 3T3 CTA report:

(f) The endpoint is focus formation (number of foci) (a potential "subjectivity" issue).

(g) The test system is well described as far as the current mechanistic understanding allows (section 3.3 pp.20-21 in the BALB/c 3T3 CTA report).

(h) The protocol is detailed and clear (section 3.6 pp.26-31 in the BALB/c 3T3 CTA report), and addressing the important issues (e.g. medium, serum) individually. The controls were specified.

(i) Acceptance criteria are listed for both cell growth and transformation assays, and are clear (section3.6.6 p.30 in the BALB/c 3T3 CTA report). Type III foci (as well as type I and II) should be better defined.

3.2 Quality of the background provided concerning the purpose of the test method

NOTE: What is the overall purpose of the test method (scientific use, regulatory application, guidelines, etc.)

The overall purpose (development of OECD guidelines) was clear, but the specific purpose of the tests was neither defined by the OECD nor by the VMT.

The purpose of the prevalidation studies is to assess whether the three CTA protocols (SHE pH6.7; SHE pH7.0 and BALB/c 3T3) have been sufficiently standardised to be transferable to other laboratories and reproducible between different laboratories and may therefore be fit for future use and to address the suitability of the three CTAs/protocols to be used as a basis for the development of OECD test guidelines.

3.3 Quality of the documentation and completeness of SOPs and prediction models

NOTE: Are the SOPs sufficiently detailed and complete? Are the prediction models sufficiently well explained to be applied in the correct manner?

The SOPs were found acceptable provided some minor revisions, including the ones recommended by the VMT.

General observations:

(a) Three SOPs with specific weaknesses and strengths (section 12.2 pp.69-88, section 12.2 pp.73-98 and section 13.2 pp.87-98 in the SHE pH6.7 CTA, SHE pH 7.0 CTA and BALB/c 3T3 CTA reports, respectively). The use of the same template for the different SOPs would have been useful for comparison and in view of developing, in case of the SHE assays, a common protocol by merging the pH6.7 and pH7.0 protocols.

(b) The terms "replicate", "independent" and "random" should be better defined.

Specific observations:

For the SHE pH6.7 and SHE pH7.0 CTA reports:

(a) The prediction model (assay assessment criteria) was described in detail for the SHE pH6.7 CTA (section 3.5.8 p.24). Why was the dose-response (trend) effect not included in 2nd bullet point, 1st item ("increased transformation frequency at at least 2 dose levels")?

(b) For the SHE pH7.0 CTA, the prediction model is poorly described in the SOP (section 6.6 p.86), while it is well described in the report text (section 3.5.8 p.25).

(c) The SOPs should specify that the cell cultures should be regularly tested for mycoplasma to exclude contamination.

(d) Adjusted target cell seeding for SHE pH 6.7 and pH 7.0 CTAs (section 2.4.2 p.77 and section 5.2.2 p.82 in the SHE pH 6.7 CTA and SHE pH7.0 CTA reports, respectively): a clear description of the procedure is needed on how to do the target cell seeding in a reproducible way.

(e) Clearly describe the definition of passage and the meaning of "early passage" (passage number) SHE cells (section 3.5.5 p.22 and p.24 in the SHE pH 6.7 CTA and SHE pH7.0 CTA reports, respectively), e.g. SHE cells in secondary or tertiary culture.

(f) Provide better guidance concerning the maximum duration of storage of cryopreserved SHE cells. For example, the SHE pH6.7 CTA report states in section 4.4. p.85, 2nd paragraph that "the storage period should not exceed 24 months". It should be explained why this period should not exceed 24 months.

SHE pH 6.7 CTA SOP:

(g) Section 2.4.4 p.78, 3rd paragraph: "10-15 % aqueous Giemsa" should be "10% aqueous Giemsa". The concentration of Giemsa was defined p.80, section 2.4.11 Fixing and staining of the colonies.

(h) Section 3.1.2 p.82: Add the criteria of normal, slightly reduced, and greatly reduced colony size and density.

SHE pH7.0 CTA SOP:

(i) Section 3. p.77, line 8, section 5.2.1. p.82, line 6, section 5.2.2 p.82, line 2 and section 5.3.3 p.82, line 1: "24 hours" should be "approximately 24 hours".

(j) It is noted that the SHE pH6.7 and the SHE pH7.0 use different fixatives (e.g. section 5.6 p.84, 4th paragraph). For the SHE pH 7.0 ethanol is being used, while for the SHE 6.7"methanol was the fixative. In view of a possible harmonisation of the SHE assay SOPs, a decision to use one of the two fixatives is recommended.

<u>For the BALB/c 3T3 CTA report:</u>

(k) The SOP is adequately described (section 3.6 pp.26-31 and section 13.2 pp.87-98).

(l) The assay assessment criteria (prediction model) was described in detail and anticipates the occurrence of inconclusive results (section 3.6.7 pp.30-31).

(m) A high dose was used for the positive control (BALB/c 3T3 CTA) which may hide variations in test performance (Tables 12 and 13 in Annex 1).

(n) Status of the cells regarding mycoplasma should be clarified (section 3.6.1 p.26).

(o) Section 3.6.1 p.26, 2nd paragraph: BALB/c 3T3 cells were "...used for the CTA within 3 to 4 passages". An explanation would be helpful as to why "within 3 to 4 passages"?

(p) Figure 31 p.67:

HRI "TA1" should be "TA2". HRI "TA2" should be "TA3".

(q) p.95, Appendix 1, Culture vessels: 100x20 mm dishes should be 90x20 mm dishes (p.91, line 24 from the bottom: Authors defined the size of dishes as 90mm).

4. Data quality

The data obtained for the SHE CTAs during this study were processed and are summarized in Annex 1 (Tables 7 to 10),

4.1 Overall quality of the evaluated data

NOTE: What is the quality of the data evaluated (testing data).

In general, the data quality was good. Acceptance criteria are broad enough to anticipate different outcomes of the assays when applied properly. Discrepancies were explained.

General observations:

It is not clear from the reports whether there were pre-defined procedures for data management/handling and whether such procedures, if in place, had been followed. This concerns mainly issues such as compliance of the generated data with the defined acceptance criteria and possible corrective measures in case the acceptance criteria were not met. There seem to be cases where the criteria were indeed not met.

Specific observations:

(a) Phthalic anhydride gave conflicting results in the transforming activity. It was shown to be positive in SHE pH6.7 CTA but negative in SHE pH7.0 CTA, which was the expected result based on previous results from the literature. A possible explanation for this discordant result was offered on p.59, 2nd paragraph in section 7.6 in the SHE pH6.7 CTA report:

"This difference could be due to pH dependent instability of phthalic anhydride in the aqueous media used in the SHE assay. According to the OECD Screening Information Data Set (SIDS) document on phthalic anhydride (April 2005), phthalic anhydride is unstable in water, hydrolyzing within minutes completely to phthalic acid which is non-genotoxic. Importantly, experiments with phthalic anhydride performed in the presence of buffer showed a half-life for phthalic anhydride of 30.5 seconds at pH 7.24 and at pH 6.8 the half-life of phthalic anhydride in water was prolonged to 61 seconds. Thus, small differences in the dose preparation, timing and pH stability of phthalic anhydride in the SHE assays could have contributed to the conflicting results obtained in this study."

The ESAC WG believes that more discussion may be necessary to clarify the different response for phthalic anhydride, because one cannot ignore the possibility that the difference may stem from the substantial difference of SHE cells cultured in low pH medium rather than a suspected immediate effect of pH on the test items (as suggested by the VMT's explanation).

(b) Retesting in case of unqualified tests

pp.65-66, module 4, section 6.5.2.3. in the BALB/c 3T3 CTA report: The VMT requested HRI to repeat the experiments with BALB/c 3T3 cells treated with o-toluidine HCl. The results obtained from the cells treated with 800 to 1200 •g/ml o-toluidine HCl were negative in Transformation Assay 2 and positive in Transformation Assay 3. The VMT concluded that treatment with o-toluidine HCl in the repeated tests of HRI and in the other studies

conducted by the other participating laboratories produced a positive response in morphologically transformed foci; however the ESAC WG questions why the VMT gave a positive decision for the results from HRI based on only two contradictory experiments. HRI should have conducted further repeat experiments until reproducible results were obtained without exceeding the number of admissible retests which should have been predefined beforehand.

The same issue was observed in the results shown in Tables 22 and 23, p.59, in which transformation assay results from ECVAM, testing phenanthrene conducted with BALB/c 3T3 are described.

More generally, the ESAC WG noted that the number of admissible retesting runs in case of unqualified tests apparently had not been defined which should ideally be done in any type of prevalidation or validation study.

4.2 Sufficiency of the evaluated data in view of the study objective

NOTE: Are the data and their quality sufficient in view of the stated objective of the study?

The data generated and evaluated did not allow for either a proper assessment of within-laboratory reproducibility or the success of the test transfer within the context of a dedicated study phase (transferability) for all three assays.

In contrast, the data produced for assessing between-laboratory reproducibility were considered sufficient for the SHE assays and may moreover be used to infer the success of transfer. The lack of appropriate testing for both modules - within-laboratory reproducibility and transferability - is, however, not considered compliant with standard practice in validation.

Specific observations:

(a) Within-laboratory reproducibility was established using a single compound, which was also the positive control. This may lead to an overestimation of reproducibility due to the compound's high intrinsic transforming activity.

(b) An experiment specifically assessing the success of the training and transfer from the lead-laboratory to the other laboratories was neither described nor performed.

(c) The description of the transfer module for all three assays is patchy: information on the timeline and the specific actions undertaken would have been helpful. In particular, more information on how the photo catalogues have been developed as a result of the "transferability module" would be helpful.

(d) For the SHE assays, considerable differences were observed between the dose-response curves produced by the different laboratories.

The ESAC WG notes that for some substances the SHE assay has been observed historically to give non-monotonic (so-called "bell shaped") curves, i.e. that the number of transformed colonies drops off when further increasing the dose. The reason for this is not fully understood, but may be caused by disturbed conditions for expression of the transformed phenotype at increased doses. Thus, there may be competition between the ability of a compound to induce transformed morphology, and the ability to inhibit the effect at higher concentrations. The sensitivity of the assay may for similar reasons vary between experiments.

Nevertheless, despite these differences, the final predictions were in most cases concordant. This shows that the assay assessment criteria are robust enough to manage such differences which may be expected in the real testing situation, when no information on dosage and increments is available.

Examples for the SHE pH6.7 CTA: MTF caused by benzo(a)pyrene (Figure 5 p.29 in the SHE pH6.7 CTA report), MTF caused by 2,4-diaminotoluene (Figure 12 p.41 in the SHE pH6.7 CTA report), MTF caused by o-toluidine HCl (Figure 18 p.50 in the SHE pH6.7 CTA report).

Examples for the SHE pH7.0 CTA: MTF caused by 2,4-diaminotoluene (Figure 8 p.43 in the SHE pH7.0 CTA report), MTF caused by o-toluidine HCl (Figure 14 p.53 in the SHE pH7.0 CTA report)

(e) For the BALB/c 3T3 assay refinement of the assessment criteria is required.

4.3 Quality of the reference data for evaluating reliability and relevance[3]

NOTE: What is the quality of the reference data used? Are the data and their quality sufficient in view of the study objective?

The quality of the reference data was assumed sufficient for assessing reproducibility, being based on the OECD DRP and well-regarded sources (i.e. IARC, Gold & Zeiger and NTP database). However, it was noted that there were apparently no provision for assessing the quality of data reported in the OECD DRP and consequently in the present study.

Specific observations:

Although not specified, the small amount of human data available means that the CTAs assessed here are used as prediction tools for rodent rather than human carcinogenicity.

[3] OECD guidance document Nr. 34 on validation defines relevance as follows: "Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of accuracy (concordance) of a test method."

5. Test materials

5.1 Sufficiency of the number of evaluated test items in view of the study objective

NOTE: Is the number of test items tested during the study sufficient in order to draw conclusions with respect to the objective of the study? If not, are there reasons for deviations and are these explained and justified?

The number of chemicals (n=6) is judged to be sufficient, with respect to statistical requirements, to assess reproducibility (the main study objective).

However, although it is acknowledged that this is not a full validation study, the number of substances tested is low. More specifically, the number appears low to adequately cover, also for the purposes of a prevalidation study, the range of possible types of chemicals in view of the most prominent underlying mechanism of action (i.e. genotoxic/non-genotoxic) for an endpoint as complex as carcinogenicity.

Thus, the reproducibility assessment is restricted in this case to substances that belong to the same chemical classes (i.e. organic substances; inorganic compounds have not been tested) and which have the same mechanism of action as the ones tested in the prevalidation study. Considering the more advanced SHE assays, 3/6 of the substances were clear genotoxic carcinogens, only 1/6 of the substances was a possible non–genotoxic carcinogen (more details in section 5.2).

Specific observations:

(a) The chemicals for assessing reproducibility were selected on the basis of available in vivo and in vitro data (see comments under section 1. Data Collection, see also table 35, p.68 in the SHE pH6.7 CTA report).

(b) The number of compounds is small and includes only compounds with historically clear positive or clear negative test results in the literature. It is debatable whether the performance of the test methods was adequately challenged by such clear transforming/non-transforming chemicals that are more likely to give concordant and hence reproducible results.

The inclusion of chemicals where there are discordant results in the literature might have provided information on the test performances where transferability or between-laboratory reproducibility could be more difficult to confirm.

(c) The number of compounds was sufficient for the main purpose of the study (reproducibility and transferability) although it is noted that transferability was not assessed directly by testing chemicals in dedicated experiments.

5.2 Representativeness of the test items with respect to applicability

NOTE: Analysis of how well the test items were selected in order to gain – through empirical testing during the study – insight into the applicability domain / limitations of the test method OR analysis to which extent the test items used during the study map an applicability domain already known.

It is not the objective of this study to assess the limitations of the tests. However, based on the ESAC WG's analysis (see section 12.1; Annex 2) showing the apparent similarity of the historical protocols compared with the protocols from this prevalidation study, it is conceivable that historical data as reported in the OECD DRP could help describe the applicability domain of the test methods in the future.

When considering the test items of this study, it appears that mainly clear positives (transforming agents) and clear negatives (non-transforming agents) have been tested, but no "equivocal" substances (known to be able to lead to discordant results within and between laboratories) which may have challenged the reproducibility of the protocols more adequately.

For the two SHE protocols, a few materials, such as reserpine, cinnamyl anthranilate, or ethylene thiourea, which have given discordant results in previous interlaboratory studies (Tu et al., 1986; Jones et al., 1988) might have helped better define the transferability and reproducibility of these newly standardised protocols.

When considering the complexity of the endpoint (i.e. regarding decisions on "genotoxic/non-genotoxic" and "carcinogenic/non-carcinogenic") as well as considering the small number of test items (n=6), the test items covered a range of the possible combinations of (non)genotoxic and (non)carcinogenic (see Figure 2 "categorisation tree" and Box 1 reproducing table 35, p.68 in the SHE pH6.7 CTA report; see also Figure 3 "categorisation tree" and Box 2 reproducing table 36, p.86 in the BALB/c 3T3 CTA report)[4].

Briefly, in case of the SHE assays, 4/6 substances tested are carcinogens. These are benzo(a)pyrene, 2,4-diaminotoluene, o-toluidine HCl and 3-methylcholantrene. 2/6 are non-carcinogens when considering reference data from the rodent bioassay (anthracene and phtalic anhydride). One of these non-carcinogens (Anthracene) is currently not classifiable according to IARC.

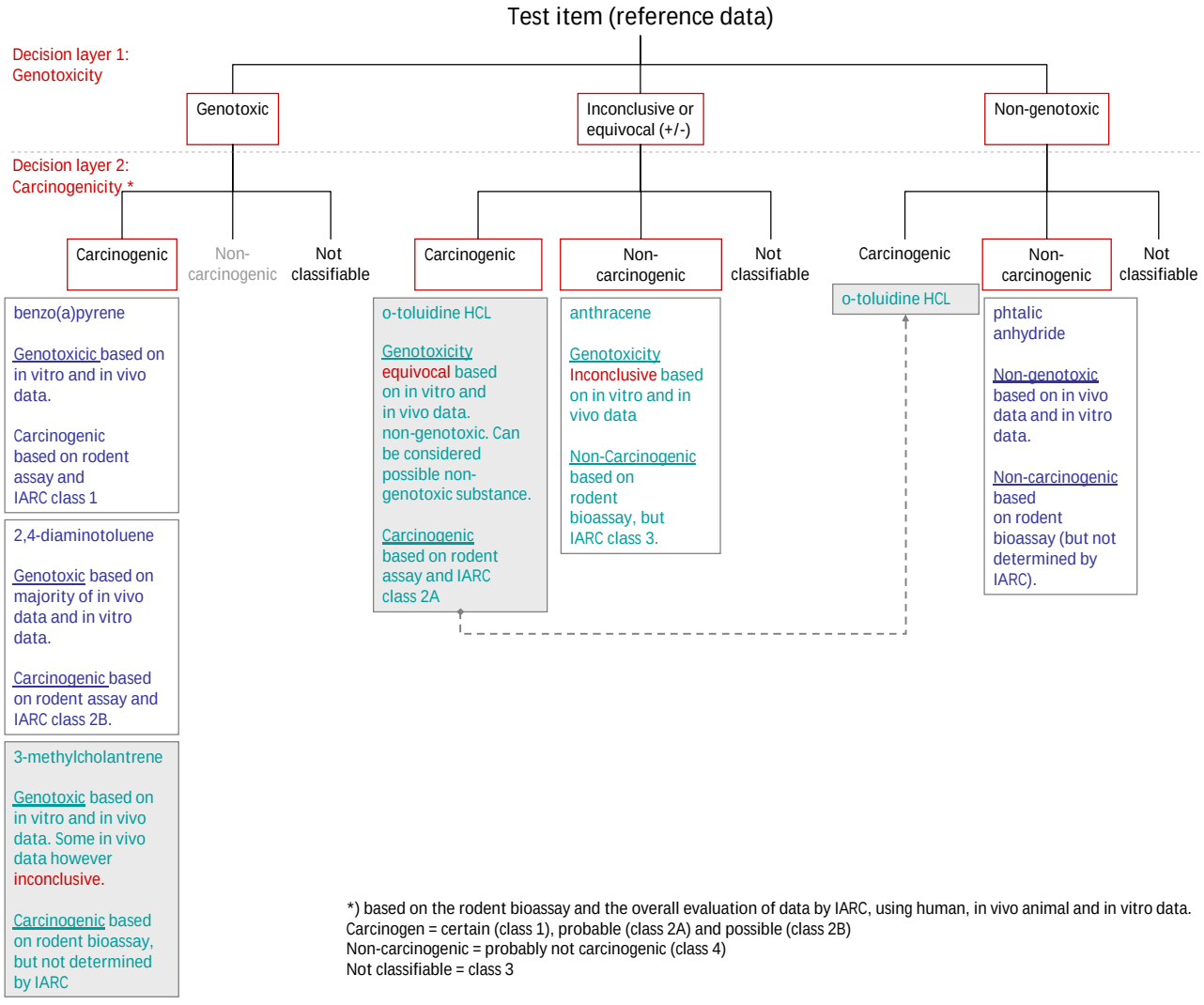
Furthermore, 2/4 carcinogenic substances studied are clearly genotoxic in in vivo and in vitro assays (benzo(a)pyrene and 2,4-diaminotoluene), while for one the overall evidence suggests that it is a genotoxic carcinogen despite some inconclusive in vivo genotoxicity data (3-methylcholantrene)[5]. The remaining substance has equivocal data from in vivo and in vitro genotoxicity tests (o-toluidine HCl).

As the concept of non-genotoxic carcinogens has only been accepted rather recently and many substances found to be carcinogens have been tested repeatedly for genotoxicity, it is possible that such substances with equivocal genotoxicity data could be regarded as non-genotoxic carcinogens (e.g. o-toluidine HCl).

[4] For an example of the aspects to be taken into consideration with regard to chemical selection for carcinogenicity/genotoxicity testing, see Vinken et al. (2008). For a full bibliographic reference see section 16.

[5] Evidence for the genotoxicity of 3-Methylcholantrene is provided for instance in: Moorthy et al., 2007; Xu et al., 2005; Rihn et al., 2000; Moorthy et al., 1993; Bryla & Wyand, 1992. For bibliographic references see section 16.

Figure 2: Categorisation tree of the SHE test items showing the possible combinations of (non)genotoxic, (non)carcinogenic and inconclusive/equivocal and the placing of the test items assessed for reproducibility within these categories. Theoretically at least 9 different categories of substances are conceivable based on the mutual combinations of (non)genotoxic, (non)carcinogenic and inconclusive/equivocal. Substances with clear conclusive data in blue [Genotoxic carcinogens: benzo(a)pyrene and 2,4-diaminotoluene; Non-genotoxic non-carcinogen: Phtalic anhydride]; substances with some degree of uncertainty regarding their belonging to one of the logical categories in pale green. Two of the four carcinogenic substances (shaded in grey) have equivocal data or some inconclusive data (highlighted in red) on genotoxicity. While o-toluidine HCl can be regarded as a possible non-genotoxic carcinogen (dashed arrow) based on its equivocal data, 3-methylcholantrene is generally seen as a genotoxic carcinogen on the basis of the overall evidence available despite some inconclusive in vivo genotoxicity data.



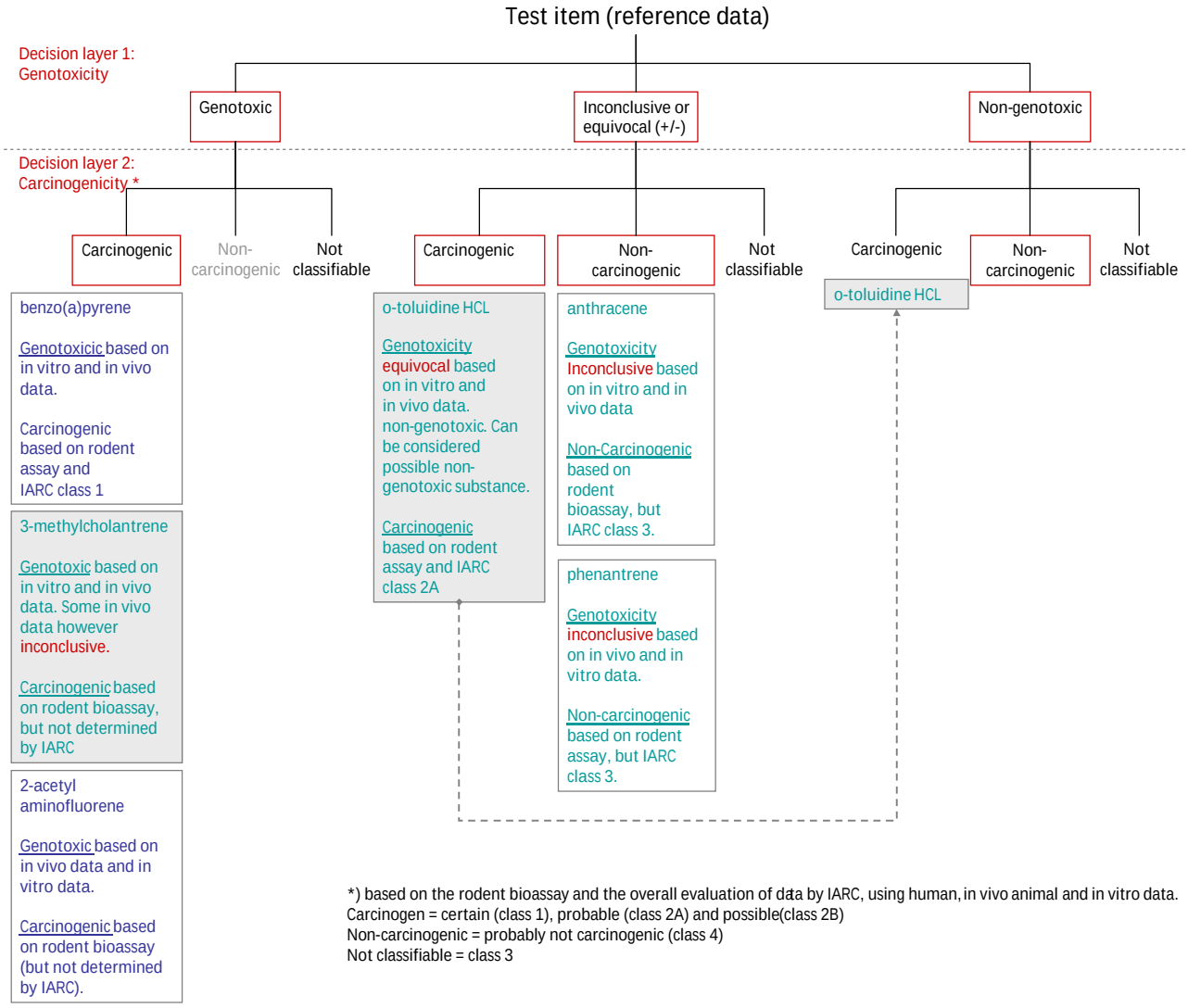
Box 1: Table 35 of the SHE pH6.7 CTA report, outlining the carcinogenicity data and genotoxicity data available for the substances tested. The data are graphically summarised in Figure 2.

Table 35: Genotoxicity and carcinogenicity data on the chemicals selected for the SHE pH 7.0 CTA prevalidation study

Chemical	CAS number	Genotoxic profile *in vitro*				Genotoxic *in vivo*	IARC class	*in vivo* carcinogenicity
		Ames	MLA	MNT	CA			
Benzo(a)pyrene	50-32-8	+	+	+	+	+ (gene mutation, MN)	1	+
Anthracene	120-12-7	+/-	+	nd	-	i	3	-
2,4-Diaminotoluene	95-80-7	+	+	nd	+	+ (UDS, transgenic mutant, comet) - (MN)	2B	+
3-Methylcholanthrene	56-49-5	+	+	+	+	i	nd	+ *
o-Toluidine HCl	636-21-5	+/-	+/-	+/-	+/-	+/-	2A	+
Phthalic anhydride	85-44-9	-	+	nd	+	- (gene mutation)	nd	- **

* source: Gold and Zeiger (1997); ** source: NTP database
+: positive; -: negative; +/-: diverging results inside a database; nd: not determined; i: inconclusive result;
MLA: Mouse Lymphoma Assay; MNT: Micronucleus Test; CA: Chromosome Aberration; UDS: unscheduled DNA synthesis; MN: micronucleus.

Figure 3: Categorisation tree of the BALB/c 3T3 test items showing the possible combinations of (non)genotoxic, (non)carcinogenic and inconclusive/equivocal and the placing of the test items assessed for reproducibility within these categories. Theoretically at least 9 different categories of substances are conceivable based on the mutual combinations of (non)genotoxic, (non)carcinogenic and inconclusive/equivocal. Substances with clear conclusive data in blue [Genotoxic carcinogens: benzo(a)pyrene and 2-acetylaminofluorene]; substances with some degree of uncertainty regarding their belonging to one of the logical categories in pale green. Two of the four carcinogenic substances (shaded in grey) have equivocal data or some inconclusive data (highlighted in red) on genotoxicity. While o-toluidine HCl can be regarded as a possible non-genotoxic carcinogen (dashed arrow) based on its equivocal data, 3-methylcholantrene is generally seen as a genotoxic carcinogen on the basis of the overall evidence available despite some inconclusive in vivo genotoxicity data.



*) based on the rodent bioassay and the overall evaluation of data by IARC, using human, in vivo animal and in vitro data.
Carcinogen = certain (class 1), probable (class 2A) and possible(class 2B)
Non-carcinogenic = probably not carcinogenic (class 4)
Not classifiable = class 3

Box 2: Table 36 of the BALB/c 3T3 CTA report, outlining the carcinogenicity data and genotoxicity data available for the substances tested. The data are graphically summarised in Figure 3.

Table 36: Genotoxicity and carcinogenicity data on the chemicals selected for the Balb/c 3T3 CTA prevalidation study

Chemical	CAS number	Genotoxic profile *in vitro*				Genotoxic *in vivo*	IARC class	*in vivo* carcinogenicity
		Ames	MLA	MNT	CA			
3-Methylcholanthrene	56-49-5	+	+	+	+	i	nd	+*
2-Acetylaminofluorene	53-96-3	+	+	+	+	+ (gene mutation, MN)	nd	+*
Benzo(a)pyrene	50-32-8	+	+	+	+	+ (gene mutation, MN)	1	+
Anthracene	120-12-7	+/-	+	nd	-	i	3	-
Phenanthrene	85-01-8	+/-	nd	nd	nd	i	3	-
o-Toluidine HCl	636-21-5	+/-	+/-	+/-	+/-	+/-	2A	+

* source: Gold and Zeiger (1997)
+: positive; -: negative; +/-: diverging results inside a database; nd: not determined; i: inconclusive result;
MLA: Mouse Lymphoma Assay; MNT: Micronucleus Test; CA: Chromosome Aberration; MN: micronucleus.

Specific observations:

The ESAC WG has the following observations concerning the selection criteria applied for identifying chemicals:

1. By requiring materials to be positive in both SHE and BALB/c 3T3 assays, one limits the number of eligible chemicals to those most widely studied. Most widely studied chemicals however are likely to be those that give clear responses. Thus, this selection criterion may introduce a bias towards good reproducibility.

   However, using identical chemicals for all three protocols may also be seen as an attempt to secure comparability of the data of the validation study which is good practice.

2. By requiring a set of chemicals negative in both SHE and BALB/c 3T3 assays, one limits the number of eligible substances again to those most widely studied, possibly introducing bias (see point 1).

   However, as stated under 1., also this criterion may be seen as good practice from the viewpoint of comparability of the data generated in the study.

3. Requiring two references for each chemical may again increase the bias towards more widely studied substances (see point 1).

4. Requiring a clear classification of the test items as an in vivo carcinogen or non-carcinogen may again lead to a bias towards selection of materials with unequivocal results and therefore to the selection of substances that will not challenge transferability or between laboratory reproducibility to the extent possible.

   However, while in vivo reference data were not necessary for the declared objective of the study (i.e. reproducibility assessment) reference data on in vivo effects allow evaluating the predictive capacity of the protocols. Furthermore, these reference data provide a reference point in case of discordant results during a (pre)validation study when assessing reproducibility. These reference data can thus facilitate the identification of possible problems of study conduct in one of the participating laboratories.

5. Both the SHE and BALB/c 3T3 assays were assessed on the basis of organic compounds only. There are a significant number of inorganic compounds in the database which could have been used to inform about reproducibility of the new protocols for inorganic substances.

   However, it is acknowledged that it is not the scope of this prevalidation study to investigate applicability or limitations of these protocols. Furthermore, the limitation to a small number of 6 test items does not allow for the covering of a broad range of chemical classes.

6. Within-laboratory reproducibility (Module 2)

6.1 Assessment of repeatability and reproducibility in the same laboratory

NOTE: How were repeatability and reproducibility assessed? Are the conclusions justified by the data as evaluated?

The ESAC WG feels that within-laboratory reproducibility was not clearly established due to inadequate study design: only one chemical was tested. Moreover, this substance was the positive control which may lead to an overestimation of reproducibility due to the clear effects to be expected.

General observations:

(a) The design for assessing the within-laboratory reproducibility was variable with regard to the different protocols assessed (e.g. coded versus non-coded for SHE pH6.7 CTA while three dosed repetitions for SHE pH7.0 CTA).

(b) Within-laboratory reproducibility was based on only one chemical which was, moreover, the positive control (producing a possible bias towards reproducible results).

Specific observations:

(a) In the SHE pH6.7 CTA, each of the three laboratories performed the assay using benzo(a)pyrene first as a non-coded and then as a coded chemical. There was no clear dose-dependent response in terms of MTF. Comparison of MTF and relative plating efficiency data within the laboratories showed a good reproducibility, although only the non-coded benzo(a)pyrene from one laboratory gave a higher MTF value, probably due to a high cytotoxicity (pp.29-30, Figure 6).

(b) In the SHE pH7.0 CTA, one laboratory performed three independent assays using coded benzo(a)pyrene. In addition, benzo(a)pyrene was used as the positive control for the assay conducted with coded chemicals. There was no clear dose-dependent increase in MTF (Table 11 in Annex 1). The results obtained by the University of Metz were quite different from those obtained by the other laboratories involved in the study.

All the experiments conducted in the same laboratory (University of Metz) were reproducible showing significant increases in MTF compared to the vehicle control at all test concentrations equal to or higher than 0.1 •g/ml (section 4.11 pp.26-27 in the SHE pH7.0 CTA report).

(c) In the BALB/c 3T3 CTA, each of the three laboratories performed the assay using 3-methylcholanthrene as a non-coded and coded chemical. There was no significant dose-dependent increase in the mean number of foci (Table 12 in Annex 1).

The trend in the formation of transformed foci by non-coded and coded 3-methylcholanthrene was similar in all laboratories.

6.2 Conclusion on within-laboratory reproducibility as assessed by the study

NOTE: How was within-laboratory reproducibility assessed? Are the conclusions justified by the data as evaluated?

The VMT's conclusion of acceptable within-laboratory reproducibility is not justified because of inadequate study design: only one chemical had been tested.

As outlined in section 6.1, only one chemical was tested to assess within-laboratory reproducibility. This chemical was the positive control (benzo(a)pyrene for SHE CTAs, 3-methylcholanthrene for the BALB/c 3T3 CTA), which may, due to its strong transforming potency, lead to an overestimation of reproducibility.

Within-laboratory reproducibility for this substance was, not surprisingly, high. However, one chemical only (being in addition the positive control) cannot be considered a sufficient dataset to conclude on within-laboratory reproducibility in compliance with what is generally considered standard practice in validation.

7. Transferability (Module 3)

7.1 Quality of design and analysis of the transfer phase

NOTE: Was the transfer phase appropriately planned, e.g. transfer instructions, training, minimum requirements, training SOP (if appropriate). Where evaluation / decision criteria defining a successful transfer established beforehand and consistently applied during the analysis?

The transfer phase was adequately described and appropriately executed so allowing proper test method conduct in the other laboratories for the subsequent analysis of between-laboratory reproducibility. However, how the success of the transfer was assessed and what criteria were used to judge the transfer successful was not clearly described. Moreover, ease of transferability was not assessed through the testing of test items. While this is not a prerequisite for prevalidation studies, the current study nevertheless did not fully address one of its objectives (i.e. assessment of the transferability module).

Specific observations:

(a) It was not described how the success of the transfer was evaluated before the assessment of the between-laboratory reproducibility. The reports suggest however that within-laboratory reproducibility data of the positive control were intended to provide evidence on transferability (see for instance section 2.4.2 p.17 and 2.6, 5th paragraph on pp.18-19 in the SHE pH7.0 CTA report).

(b) Photo catalogues were established for assuring consistency for the assessment of the morphology of transformed colonies (for the SHE assays) and transformed foci (for the BALB/c 3T3 assay) during the scoring of the experiments performed to assess the between-laboratory reproducibility

(c) Representatives from all laboratories involved (including technical staff and study directors) participated in a training for harmonizing the procedures across the laboratories for the prevalidation studies.

7.2 Conclusion on transferability to a second laboratory as assessed by the study

NOTE: Are the conclusions justified by the data generated? Have critical issues that may impact on transferability been identified?

The success of the transfer programme was not demonstrated in separate experiments. All participating laboratories had some experience with CTAs. Thus, the ease of transferability to a laboratory without any CTA experience was not demonstrated. However, this is in any case not a formal requirement for a prevalidation study (OECD guidance document Nr. 34), and in this case it is noted that successful transfer may be inferred from the good between-laboratory reproducibility (see below).

The study design was appropriate for ensuring transfer of the protocols to other laboratories and the generation of photo catalogues likely to support a more consistent approach to visual scoring is considered one of the merits of the study.

In contrast, the ease of transferability to naive/inexperienced laboratories was not assessed through empirical testing of, at least a subset, of the test items, contrary to reports which state that one of the objectives was to "assess" the transferability module (e.g. p.9, 2nd paragraph in the SHE pH7.0 CTA report). While ease of transferability is not a specific requirement of prevalidation studies (OECD Guidance Document Nr. 34)[6], it may be considered good practice in validation to include such a testing phase to give information about the possible problems that may arise in transferring a method and/or its robustness in view of possible minor (uncontrolled) alterations of test procedure. This would have been important in this specific case, since the test method relies on an observational readout (visual scoring) from which issues of transferability are likely to arise. However, since ease of transferability is not a formal requirement for such a study, lack of testing is not considered crucial for the conclusions of the study.

Moreover, it should not be omitted that the VMT has obviously given this issue some thought and the ESAC WG acknowledges that the reports discuss, for example, to which extent within-laboratory variability data of the different laboratories (and all based on testing the positive control only) could be useful in making conclusions on ease of transferability. See for instance sections 2.4.2 p.17 and 2.6, 5th paragraph on pp.18-19 in the SHE pH7.0 CTA report, the latter reproduced below:

"Following the preliminary phase of optimisation of the protocols, the transferability and the within-laboratory reproducibility were assessed by evaluating results obtained for one non-coded test chemical and a coded one. These two chemicals were the same (benzo(a)pyrene), allowing an analysis of the within-laboratory reproducibility as well as the transferability of the assay."

[6] The guidance document outlines that "Additional activities at the prevalidation stage may include an assessment of the transferability of the test protocol to laboratories inexperienced in the test or the necessary techniques and to resolve questions or inconsistencies arising in this phase" (Paragraph 75, OECD Guidance Document Nr. 34 on the Validation and International Acceptance of new or updated test methods for hazard assessment, OECD 2005)

8. Between-laboratory reproducibility (Module 4)

8.1 Assessment of reproducibility in different laboratories

NOTE: How was reproducibility between laboratories assessed?

The final outcome following the implementation of the assessment criteria was considered reproducible for the SHE assays. For the BALB/c 3T3 assay, refinement of the assessment criteria is required.

General observations:

(a) The dose-ranges differed between laboratories. The VMT had decided to indicate clear dose ranges and increments for some compounds, while for others no clear indications of the dose points and increments had been given (see citation from Validation Study Report below). The latter case was intended to assess performance of dose-range finding in a realistic testing situation, when dose range and increments are not clear. This provides information on the potential transferability of the test methods.

In the case of the SHE pH6.7 CTA the VMT report states (p.17, paragraph below table 1):

"The doses of benzo(a)pyrene, 3-methylcholanthrene and o-toluidine HCl to be used were suggested by the VMT based on data from the literature to optimise the use of resources (either due to high chemical cost or lack of cytotoxicity) for timely completion of these studies. For the other chemicals the laboratories were asked to select the dose ranges on their own in order to check their ability to identify the critical doses for the transformation assay."

(b) In the context of the dose range finding, the ESAC WG recommends to include guidance for dose selection in the SHE protocols to ensure proper use of the protocols in testing laboratories. The following approach is suggested:

The highest dose should be determined base on relative plating efficiency and the lower doses spaced out for instance with two concentrations per log (for instance: 100, 30, 10, 3, 0.3, 0.1 etc.).

(c) The doses may have to be spaced differently for the BALB/c 3T3 assay since the responsive concentrations are spaced in a narrower band as compared to the SHE. This could be the reason why repetitions of BALB/c 3T3 CTA experiments with more narrow doses resulted in change from 'negative' to 'positive' calls/predictions.

Based on the study data, there is therefore a need to optimize the number and spacing of doses in the BALB/c 3T3 CTA SOP (e.g. 8 doses with what spacing?). For instance, in the BALB/c 3T3 CTA report some of the dose-responses are quite steep and could be missed if the dose spacing is not correct.

Specific observations:

For the SHE pH6.7 and SHE pH7.0 CTA reports:

(a) Reproducibility is acceptable.

(d) It is not clear why the concentration of benzo(a)pyrene for the positive control differed between laboratories.

(b) SHE pH6.7 CTA: It is standard practice to repeat inconclusive results (e.g. o-toluidine HCl), however, the number of retesting runs should be defined in a study and, moreover, guidance should be provided in the SOP.

(c) SHE pH7.0 CTA: The phthalic anhydride result produced by BASF is different from the results produced by the other laboratories. Although the experiment met the assay assessment criteria (specified for testing in one laboratory), it would have been perhaps advisable to repeat this experiment since discordant results had been produced when considering the results of all laboratories.

(d) After implementation of the acceptance criteria, reproducibility is good.

For the BALB/c 3T3 CTA report:

(e) The mean numbers of foci per dish in each of the three laboratories were higher than those observed during the within-laboratory reproducibility assessment (Table 13 in Annex 1).

(f) It is not clear why incorrect solvents were used by one laboratory (e.g. DMEM-L and water rather than DMSO). It is not clear whether this indicates (1) a lack of appreciation of the study objective, (2) errors at a technical level or (3) disagreement followed by a unilateral choice of what was believed to be the best method.

(g) Only three (out of six) chemicals produced concordant results across all laboratories before the repetition of some experiments.

(h) The ESAC WG agreed with the conclusion of the VMT that the between-laboratory reproducibility was not satisfactory before retesting and improved after retesting.

8.2 Conclusion on reproducibility as assessed by the study

NOTE: Are the conclusions justified by the data generated?

Between-laboratory reproducibility was assessed through analysis of the concordance of predictions for the six test substances obtained by the involved laboratories. The predictions concerned the classification of test substances as potential transforming agents/non-transforming agents in the CTA assays. The CTA predictions were compared with the reference data associated with the test chemicals. These data are in vivo carcinogenicity predictions taken from the OECD DRP report which, for the test chemicals, are based on IARC classifications, the Gold & Zeiger and the NTP databases.

Based on the data generated and reported, the ESAC WG agrees with the VMT that the two SHE protocols yield results which are concordant between laboratories and hence reproducible for the substances tested.

paragraph continued on next page

In contrast, evidence supporting reproducibility of the results between laboratories for the BALB/c 3T3 protocol was considered insufficient, as suggested by the need to refine assay assessment criteria and to repeat some experiments to obtain concordant results across the laboratories.

9. Predictive capacity (Module 5)

Although predictive capacity was outside the scope of the study objective, it is noteworthy that the predictions made by the SHE assays for the six chemicals were in most cases correct (6/6 predictions were correct in the SHE pH7.0 CTA study, while 5/6 were correct in the SHE pH6.7 CTA study). While the chemicals selected may have a bias towards reproducible results (clear negatives and strong positives), the results are nevertheless reassuring and add to the database of CTA testing data.

However, based on the ESAC WG's analysis (section 12.1; Annex 2) showing apparent similarity of the historical protocols with the protocols from the prevalidation study, it is conceivable that historical data as reported in the OECD DRP could help describe the predictive capacity of the SHE CTA test methods in the future.

10. Applicability domain (Module 6)

Since this study is not a full validation study, the assessment of the applicability domain is rather limited.

However, based on the ESAC WG's analysis (section 12.1; Annex 2) showing apparent similarity of the historical protocols with the protocols from the prevalidation study, it is conceivable that historical data as reported in the OECD DRP could help describe the applicability and limitations of the SHE CTA test methods in the future.

11. Performance standards (Module 7)

Not applicable to the current CTA study.

12. Readiness for standardised use

12.1 Assessment of the readiness for regulatory purposes

NOTE: Is the test method ready for regulatory purposes? If yes, why? If no – what impediments currently exclude application for regulatory purposes?

The data generated during this prevalidation study, when viewed on their own, are insufficient to draw conclusions on readiness for regulatory use of the SHE assay protocols, although these showed acceptable reproducibility.

However, an analysis by the ESAC WG identified considerable similarity in the historical SHE protocols and the standardised protocols in this prevalidation study. This supports the view that a substantial amount of the existing testing information from the SHE assays could be used for future considerations on their performance (e.g. predictive capacity, applicability/limitations) required to define their regulatory utility.

This view is further supported by the apparent robustness of the SHE assays as demonstrated in the OECD DRP: the data show considerable concordance with regard to the predictions made even though these older protocols may have differed to some extent and clearly no standardised test procedures had been used. These predictions are moreover relevant when compared with in vivo carcinogenicity data derived from well-respected sources (e.g. IARC, NTP database).

The ESAC WG, therefore, believes that future activities aimed at more precise definition of test method performance of the SHE assays and possible regulatory utility of the associated SHE protocols can be based on both, prospective testing but also on the analysis of existing historical information (e.g. a meta-analysis using defined search and data selection criteria based on study quality).

The ESAC WG notes, based upon current opinion, that no single method can provide sufficient information for an unequivocal assessment of the carcinogenicity potential of a substance to satisfy regulatory requirements fully. The SHE assays may provide information about possible genotoxic and non-genotoxic carcinogens for use in conjunction with other data (e.g. in the context of a "weight-of-evidence" approach). Some recommendations on possible approaches towards the expansion of the performance characterisation of these methods are made in section 15, notwithstanding the fact that the specific regulatory use needs to be defined by the relevant authorities for the purpose in mind.

The study results show that, in contrast to the SHE data, the BALB/c 3T3 protocol still requires optimisation (concerning for example the assessment criteria for the assay) and is at present neither ready to enter full validation nor consideration for regulatory use based on existing information.

Typically an assay can be considered for regulatory purposes if the three following items are sufficiently well described: (1) the toxicological effect the assay is intended to predict (its relevance), (2) how well the assay does predict that endpoint (its predictive capacity) or measure relevant mechanisms contributing to this toxicity effect and (3) how reproducible the assay is (its reliability).

The regulatory authority, however, makes the final decision of whether the level of relevance, predictive capacity and reproducibility are "good enough" for the specific purpose envisioned. On the background of these reflections, the ESAC WG holds the following views regarding the readiness of the SHE and BALB/c 3T3 assays for regulatory purposes:

The ESAC WG is of the opinion that the reliability of the BALB/c 3T3 protocol was not adequately addressed in this study. For example, repeat testing was executed without blinding and also the assay assessment criteria require further refinement. The ESAC WG understands why further refinements were made to the BALB/c 3T3 protocol, but believes that these modifications should be tested in further trials.

In contrast, acceptable reliability has been shown for the two SHE assay protocols for the compounds that were tested. Although the chemical domain of this set of compounds and the number of compounds is very limited, the ESAC WG feels that it is both likely and plausible that this level of reproducibility would extend to other chemical domains as well. Therefore, the SHE assay protocols as standardised during this study are at least sufficiently reproducible, for those chemicals tested, to be considered for eventual use in a regulatory setting.

Nevertheless, robust conclusions on the readiness for regulatory purposes of the SHE assays cannot be drawn on the basis of this prevalidation study alone. The dataset generated during this study is too small to allow sufficient characterisation of key items of test method performance. These include: reproducibility on the basis of a larger and different set of chemicals (including weakly transforming agents), predictive capacity, applicability and possible limitations.

Therefore, the next step following on from this prevalidation exercise would normally be to plan a full prospective validation study. This would comprise a set of test substances which cover a wide range of chemical classes / possible mechanisms of action (e.g. genotoxic/non-genotoxic) and is large enough for a statistical evaluation of predictions into two (dichotomous) classes: transforming or non-transforming agents.

However, it is conceivable that historical SHE testing data can be used with to help get robust performance characteristics of the SHE tests to support possible regulatory use. This is based, firstly, on the apparent robustness of the SHE assays as demonstrated by an analysis of published data in the OECD DRP (predictions were obtained using non-standardised protocols and showed nevertheless a high degree of concordance) and, secondly, the appreciable similarity of the historical SHE protocols and the standardised protocols in this study shown by the analysis carried out by the ESAC WG (see Annex 2). It should also be noted that, about 500 coded and un-coded compounds have been tested up to now using the SHE assay by many laboratories. Consideration should be given to what extent these historical data, which include validation studies, could supplement or substitute for a new full prospective validation study of the SHE assay. The ESAC WG draws attention to two publications that highlight the good predictive capacity of the SHE assays (Isfort et al., 1996; Mauthe et al., 2001).

The ESAC WG also has the opinion, that any further activities should pay attention to the chemical selection and, if using existing information, to an appropriate description of the toxicity potency of substances and their physicochemical properties and chemical class. The ESAC WG has two concerns with respect to the information compiled in the OECD DRP: (a) the lack of explicitness regarding the completeness of the existing data presented and whether data selection criteria based on study quality were defined and had been applied and (b) the lack of a description of the transforming potency of the chemicals analysed. If most data are based on studies on strong transforming agents or clear non-transforming agents (i.e. substances with no record of discordant results), then reproducibility could have been overestimated. Future activities using this information should, thus, analyse these issues before concluding on test performance and/or the planning of new prospective studies.

Specific observations:

(a) This ESAC WG considers that its responsibility is to state whether or not the studies conducted by the VMT have demonstrated sufficient reliability of the CTAs so that regulatory authorities can then make a decision of whether that level of reproducibility is “good enough” for the regulatory purpose envisioned.

(b) The ESAC WG feels that acceptable reliability has been shown for the two SHE assay protocols for the compounds that were tested. Although the chemical domain of this set of compounds is very limited, the ESAC WG is of the opinion that it is likely that this level of reproducibility extends to other chemical domains as well. The predictive capacity of the SHE assay was only addressed in a very limited fashion in the studies that we were asked to review.

(c) The ESAC WG concludes that the reliability of the BALB/c 3T3 protocol was not adequately addressed in this study because several repeat assays had to be conducted and that the repeat assays were conducted without blinding. The ESAC WG understands why further refinements were made to the BALB/c 3T3 CTA protocol, but believes that these modifications should be tested in further trials.

(d) None of the tests are currently ready as stand-alone methods for regulatory purpose. However, it is the view of the ESAC WG that they may provide useful supplementary information relevant for instance to a "weight of evidence" approach. Positive results may have more "weight" than negative ones. It should moreover be considered that due to the complexity of carcinogenicity, no single method is sufficient for a complete hazard or risk assessment.

(e) Of the different alterations introduced in historical SHE CTA protocols, the change in pH of the medium is by far the most influential factor on SHE cell morphological transformation. Nevertheless, historical data using protocols with different pH show reasonable concordance of the results obtained.

For example, the OECD DRP provides an analysis of the predictions yielded for 48 substances tested in various protocols that varied with regard to the pH of the medium (table 1-3 of the OECD DRP). The pH values were pH 6.7 and pH •7.

40 of these 48 substances (83%) gave the same response with both protocols, while 8/48 (17%) gave discordant results. Of these 8 compounds, 3 were positive (=transforming) at pH6.7 and 5 at pH •7. It is noteworthy that 4 of these 8 substances were in the category "non-carcinogenic and inconclusive for carcinogenicity" when considering the reference data (in vivo carcinogenicity predictions).

Taken together, this supports the view that historical alterations in SHE CTA protocols over time have had little impact on the reported results of chemicals tested (See section 16).

12.2. Assessment of the readiness for other uses

NOTE: Is the test method ready for other uses (e.g. screening purposes, testing to gain mechanistic insight, to generate supportive information for hazard/risk assessment).

The ESAC WG considers the CTAs useful for testing compounds belonging to the same class of chemicals as those used in the reported prevalidation studies (screening purposes) and to generate supporting information for hazard identification and risk assessment (weight of evidence). Moreover, the CTAs will continue to be useful also for mechanistic studies of the transformation process.

12.3 Critical aspects impacting on standardised use

Note: What are the factors that may impact on standardised use (in regulatory or non-regulatory settings)?

The performance characteristics of the SHE CTA methods need to be carefully analysed through prospective testing and/or analysis of existing information (protocol similarity supports the use of historical data) before the SHE CTA protocols can be used in standardised applications (regulatory or non-regulatory). This analysis should include a careful examination of the chemical classes tested. Moreover, some improvement of the SHE CTA protocols such as the development of common protocol for the two pH variants and a better description of some of the protocol steps (cell preparation) should be performed before standardised use is considered.

Concerning the BALB/c 3T3 CTA, further optimisation of the protocol is needed. These modifications, including those suggested by the VMT, should be tested in further trials before standardised use is considered.

Specific observations:

(a) Test protocols were harmonised and standardised in order to provide a basis for the development of CTA OECD test guidelines.

(b) The visual scoring of the colonies/foci had been considered one of the greatest weaknesses of the SHE and BALB/c 3T3 CTAs. Training in the conduct and scoring the colonies/foci described in the CTA reports is very likely to have made the CTAs more objective and hence more reproducible. The production of the photo catalogues is a commendable achievement of the study.

(c) Implementation of minor modifications (e.g. concerning dose selection) and alignment of the two SHE protocols are required.

(d) Further development of BALB/c 3T3 CTA protocol is required (e.g. to include proper guidance on how to approach dose spacing).

(e) Several methods to automate the scoring of morphological transformation of SHE cells have been reported on (Trevisan et al., 2010; Walsh et al., 2009; Ridder et al., 1997). It is reasonable to assume that increased recognition, acceptance and use of the assay will provide the necessary incentives to speed up this work, which could result in a low cost, objective and high throughput assay. This is especially important at a time with increased

needs for testing (e.g. the REACH challenge) and increasing restrictions on the use of animal bioassays.

12.4 Gap analysis

NOTE: Identify, if appropriate, gaps in the study design and/or execution that impact on the stated study objective or the conclusions drawn.

Specific observations:

(a) Some gaps in the study design and execution were found. These include insufficient data for within-laboratory reproducibility, no dedicated testing for transferability (although it is acknowledged that this is not always required in a prevalidation study) and issues with retesting (in particular for the BALB/c 3T3 CTA).

(b) It is not known yet how assays using these protocols will perform with equivocal (e.g. weak positive) chemicals.

(c) Procedure for data management (separated from statistician) should be established.

(d) Training programme: Criteria for success and mechanisms for assessing success were not described.

(e) It would have been appropriate, if the VMT had presented the extent to which the standardised SHE CTA protocols differed from those used historically. This would support the use of existing information when considering test method performance on a more general level.

13. Other considerations

NOTE: Please address any other consideration you might have in relation to the proposed approach under this section.

Concerning the conduct of the validation study and its conclusions:

All protocols:

(a) The high concentrations used for the positive control may hide reduced sensitivity during test performance. The concentrations should thus be reconsidered.

SHE pH6.7 CTA and SHE pH7.0 CTA:

(b) The following statement concerning the SHE pH6.7 CTA could not be confirmed by the ESAC WG (section 3.4 p.21 in the SHE pH6.7 CTA report): "Optimised cell growth reduces variation…, and also increases the number of transformants in control and treated cultures which allow for the application of robust statistical methods".

The ESAC WG carefully re-analysed MTFs of vehicle control and positive control between SHE pH6.7 and pH7.0 assays performed in the present prevalidation studies (Annex 1). There were apparently no significant differences in the MTFs of vehicle control and positive control between the cells (Tables 7-9 in Annex 1).

(c) It was noted that some dose response curves (SHE CTAs) were flat (it is conceivable that there was saturation) so not permitting the dose responses of the test to be addressed adequately.

BALB/c 3T3 CTA:

(d) Laboratory did not always follow instruction (e.g. wrong vehicle used).

(e) The criteria for an inconclusive result need strengthening.

(f) An appreciable number of experiments were repeated although the assay assessment criteria had been met. These decisions were taken on the basis of expert judgement and retesting led to changed predictions.

For example: Transformation Assay No. 1 for phenanthrene and Transformation Assay No. 2 for o-toluidine HCl were repeated due to an increase in the number of foci at the highest concentration although the results were considered negative according to the assessment criteria (see table 34 p.72 in the BALB/c 3T3 CTA report). The use of more closely spaced concentrations in the repeated experiments (Table 23 p.58 and Table 31 p.66 for phenanthrene and o-toluididne HCl, respectively, in the BALB/c 3T3 CTA report) allowed for the confirmation of the transforming effect of these two chemicals. The choice of dose levels to be tested in the transformation assay is critical to get a reliable result and should be carefully made by narrowing the inter-dose spacing, especially with chemicals displaying very steep cytotoxicity curves, as discussed by the VMT (section 8 pp.74-75 in the BALB/c 3T3 CTA report).

(g) Phenathrene was found a positive transforming agent which is, according to the VMT report "in contrast with previously published results" (section 7.5, p73). This has implications for testing of unknown substances. The ESAC WG is concerned that the use of the BALB/c 3T3

CTA protocol as evaluated in this study may result in an appreciable number of false/misleading positives.

General considerations concerning the assays:

All protocols:

(a) Despite the development of the photo catalogues supporting consistent scoring, there remains a degree of subjectivity in the assessment of colonies. The ESAC WG suggests to consider automated approaches (e.g. image analysis) as possibly more objective ways to score.

(b) It should be considered whether the SOPs should provide instructions regarding the coding of test dishes to improve objectivity of the scoring process (including positive control dishes)

(c) There is some concern that despite the development of standardised protocols as a result of this study, variants of CTA protocols may still be used. It is therefore recommended that the protocols developed here (pending some improvements, see recommended changes of SOPs in section 15) be accepted world-wide. Some attention should be paid to possible implications of results obtained previously using different protocols, in particular with regard to those protocol changes which are more likely to have an impact on the test result.

(d) The argument made in the prevalidation study reports that the lack of knowledge about the specific mechanism of action is an advantage seems a little weak. Although tests that empirically seem to work may be very useful, there is ultimately a need for tests based upon mechanisms.

BALB/c 3T3 CTA:

(e) The criteria for a positive/negative decision are not completely defined. It is not clear whether these should be solely based on a decision rule using a currently un-refereed statistical method. The ESAC WG suggests to consider the development of a decision chart which may aid consistent and transparent decision-making.

(f) The ESAC WG is concerned about the implication of the comment that "experiments requiring repetition because a biological effect was observed although the statistical analysis was negative": This suggests that the design of the statistical evaluation as described in the SOP may not be sufficient to detect effects of size that are considered biologically important. More work may need to be done on the SOP to overcome this problem.

(g) The ESAC WG is concerned about the apparent correlation between cytotoxicity observed with positive and negative compounds and the development of foci.

14. Conclusions on the study

NOTE: This section should presents an overview over the study results and conclusions as described in the study reports (subsection 14.1), discuss to which extent the conclusions drawn in the study reports are justified by the study results on their own (subsection 14.2) and evaluate to which extent the conclusions are plausible with respect to other information (subsection 14.3).

14.1 Summary of the results and conclusions of the study

The ESAC WG's description of the results of the study is summarized in Table 2 in this section. A detailed recalculation of the study results has been performed by the ESAC WG and is summarised in Annex 1. Table 3 in section 14.2 shows the extent to which the information requirements have been satisfied by the study results, taking the study type and study objectives into account. The summary results as presented by the VMT are reproduced in Annex 3.

14.1.1 Test items

The 6 items tested in the SHE CTAs were:

1) benzo(a)pyrene
2) 2,4-diaminotoluene
3) o-toluidine HCl
4) 3-methylcholantrene
5) anthracene
6) phthalic anhydride

The 6 items tested in the BALB/c 3T3 CTA were:

1) 3-methylcholantrene
2) 2-acetyl aminofluorene
3) benzo(a)pyrene
4) anthracene
5) phenanthrene
6) o-toluidine HCl

Table 1 shows to which extent the same test items were used for the assessment of the SHE and BALB/c CTAs and which were tested either only in the SHE CTAs or only in the BALB/c 3T3 CTA. The criteria used to select test items had been defined in a way to ensure greatest possible overlap between test items used to assess the three protocol variants (for a critical reflection on the chemical selection criteria see section 5.2).

Table 1: List of specific and common test items used for assessing the three CTA variants. Two substances were tested only in the SHE CTAs, two only in the BALB/c CTA, while four substances were used in all protocol variants.

Nr.	Test items used in the study (n=8)	Tested in SHE CTAs (n=6)	Tested in BALB/c CTA (n=6)	OVERLAP: tested in SHE CTAs AND BALB/c 3T3 CTA
1	benzo(a)pyrene	TESTED (PC for SHE CTAs)	TESTED	YES
2	2,4-diaminotoluene	TESTED	NOT TESTED	NO
3	o-toluidine HCl	TESTED	TESTED	YES
4	3-methylcholantrene	TESTED	TESTED (PC for BALB/c)	YES
5	anthracene	TESTED	TESTED	YES
6	phthalic anhydride	TESTED	NOT TESTED	NO
7	2-acetylaminofluorene	NOT TESTED	TESTED	NO
8	phenanthrene	NOT TESTED	TESTED	NO

The test items used for assessing the SHE protocols covered a range of the possible combinations of (non)genotoxic and (non)carcinogenic (see section 5: Figure 2 "categorisation tree" and Box 1 reproducing table 35, p.68 in the SHE pH6.7 CTA report; also Figure 3 "categorisation tree" and Box 2 reproducing table 36, p.86 in the BALB/c 3T3 CTA report), in particular, when considering the complexity of the endpoint (i.e. regarding decisions on "genotoxic/non-genotoxic" and "carcinogenic/non-carcinogenic") as well as considering the small number of test items (n=6).

The classification of test substances used for the SHE CTAs and BALB/c CTA is shown in Box 3 and Box 4, respectively.

Box 3: Categorisation of test items used in the SHE CTAs with respect to their (non)genotoxic and (non)carcinogenic profile:

A) CARCINOGENICITY[7]

A1. Carcinogenic substances:

4/6 substances tested are carcinogenic:

1) Benzo(a)pyrene
2) 2,4-Diaminotoluene
3) o-Toluidine HCl
4) 3-Methylcholantrene

A2. Non-carcinogenic substances:

2/6 are non-carcinogens when considering reference data from the rodent bioassay. 1 of these non-carcinogens (Anthracene) is currently not classifiable according to IARC.

1) Anthracene
2) Phtalic anhydride

B) GENOTOXICITY[8]

B1. Genotoxicity of the carcinogenic substances tested:

B1.1 Genotoxic carcinogens

2/4 carcinogenic substances studied are clearly genotoxic in vivo and in vitro assays:

1) Benzo(a)pyrene
2) 2,4-Diaminotoluene

For 1/4 the overall evidence suggests that it is a genotoxic carcinogen despite some inconclusive in vivo genotoxicity data:

3-Methylcholantrene

B1.2 Non-genotoxic carcinogens

The remaining substance has equivocal data from in vivo and in vitro genotoxicity tests and may be regarded non-genotoxic[9].

o-Toluidine HCl

B2. Genotoxicity of the non-carcinogenic substances tested:

B2.1 Non-genotoxic non-carcinogens

1/2 of the non-carcinogenic substances tested is non-genotoxic based on in vitro and in vivo data.

Phtalic anhydride

B2.2 Non-carcinogens with inconclusive genotoxicity results:

1/2 of the non-carcinogenic substances tested has inconclusive data from in vivo and in vitro tests:

Anthracene

[7] The conceptual term carcinogenicity encompasses both carcinogenic and non-carcinogenic substances.
[8] The conceptual term genotoxicity encompasses both genotoxic and non-genotoxic substances.
[9] As the concept of non-genotoxic carcinogens has only been accepted rather recently and many substances found to be carcinogens have been tested repeatedly for genotoxicity, it is possible that such substances with equivocal genotoxicity data could be regarded as non-genotoxic carcinogens (e.g. o-toluidine HCl).

Box 4: Categorisation of test items used in the BALB/c 3T3 CTA with respect to their (non)genotoxic and (non)carcinogenic profile:

A) CARCINOGENICITY[10]

A1. Carcinogenic substances:

4/6 substances tested are carcinogenic:

1) 3-Methylcholantrene
2) 2-Acetylaminofluorene
3) Benzo(a)pyrene
4) o-Toluidine HCl

A2. Non-carcinogenic substances:

2/6 are non-carcinogens when considering reference data from the rodent bioassay. Both of these non-carcinogens are currently not classifiable according to IARC.

1) Anthracene
2) Phenanthrene

B) GENOTOXICITY[11]

B1. Genotoxicity of the carcinogenic substances tested:

B1.1 Genotoxic carcinogens

2/4 carcinogenic substances studied are clearly genotoxic in vivo and in vitro assays:

1) 2-Acetylaminofluorene
2) Benzo(a)pyrene

For 1/4 the overall evidence suggests that it is a genotoxic carcinogen despite some inconclusive in vivo genotoxicity data:

3-Methylcholantrene

B1.2 Non-genotoxic carcinogens

The remaining substance has equivocal data from in vivo and in vitro genotoxicity tests and may be regarded non-genotoxic[12].

o-Toluidine HCl

B2. Genotoxicity of the non-carcinogenic substances tested:

B2.1 Non-carcinogens with inconclusive genotoxicity results:

2/2 of the non-carcinogenic substances tested have inconclusive data from in vivo and in vitro tests:

Anthracene
Phenanthrene

[10] The conceptual term carcinogenicity encompasses both carcinogenic and non-carcinogenic substances.
[11] The conceptual term genotoxicity encompasses both genotoxic and non-genotoxic substances.
[12] As the concept of non-genotoxic carcinogens has only been accepted rather recently and many substances found to be carcinogens have been tested repeatedly for genotoxicity, it is possible that such substances with equivocal genotoxicity data could be regarded as non-genotoxic carcinogens (e.g. o-toluidine HCl).

14.1.2 Summary of study results

Study results are summarised below. Table 2 (overleaf) presents graphically the results with respect to the study objective.

For the SHE pH6.7 and SHE pH7.0 CTA reports:

(a) The study results show that sufficiently standardised protocols are available which appears to be transferable (at least to experienced laboratories) and which are reproducible between laboratories on the basis of the six chemical tested.

(b) Moreover, as result of the transfer phase, a photo catalogue supporting the consistent scoring of visually assessed effects (number of transformed colonies) has been produced. It is conceivable that this catalogue may have supported between laboratory reproducibility, although this has not been tested (i.e. by including another laboratory working without photo catalogue). Clearly, colony scoring appeared to be less of an issue as compared with concerns raised in the past when no photo catalogue was available.

For the BALB/c 3T3 CTA report:

(c) An improved BALB/c protocol has been developed as a result of this prevalidation study and recommendations of the VMT made during the course of this prevalidation study. The recommended changes related to data interpretation only (refinement of acceptance and assessment criteria) and not to any procedural aspects related to the practical execution of the test method. However, more experimental data are required to further refine the BALB/c 3T3 CTA protocol and assess its within- and between-laboratory reproducibility and its transferability.

(d) In particular the statistical methods used requires further attention, better definition and refinement of the decision criteria based upon them. Considerations concerning the statistical methods may affect other aspects of the design of possible future prospective studies (number of plates, concentrations tested, requirements for repeat studies, etc).

(e) Moreover, as result of the transfer phase, a photo catalogue supporting the consistent scoring of visually assessed effects has been produced. With the appropriate training and the use of the photo catalogue, the scoring of foci was not problematic despite the concerns raised in the past.

Table 2: The overview to which extent the information requirements were addressed/assessed in view of the objective of the prevalidation study:

CTA variant	Test definition / Protocol standardisation	Within laboratory variability	Transferability	Between laboratory variability
SHE pH 6.7	Scientific basis of the test defined SOP and photo catalogue produced	One chemical (positive control benzo(a)pyrene) tested coded and non-coded, in 1 laboratory (for all laboratories) Same prediction (call) and similar morphological transformation frequency induced for coded and non-coded chemical	No dedicated testing Laboratory training organised Photo catalogue produced during the training to support consistent scoring	Six chemicals (4 carcinogens, 2 non-carcinogens) tested in 3 laboratories 6/6 chemicals were concordantly identified by the laboratories (but one non-carcinogenic chemical phthalic anhydride consistently identified as carcinogen by all laboratories). Although the SHE pH6.7 showed higher concordance than the pH7.0 variant, it should be noted that one of the concordant predictions was incorrect with respect to the reference data.
SHE pH 7.0	Scientific basis of the test defined SOP and photo catalogue produced	One chemical (positive control benzo(a)pyrene) tested coded in 1 laboratory with 3 repeats Same prediction (call) and similar dose-dependent induction of morphological transformation frequency for all 3 repeats	No dedicated testing Laboratory training organised Photo catalogue produced during the training to support consistent scoring	Six chemicals (4 carcinogens, 2 non-carcinogens) tested in 4 laboratories 5/6 chemicals were concordantly identified by the laboratories (but one non-carcinogenic chemical phthalic anhydride identified as carcinogen by one laboratory). Although the SHE pH7.0 showed lower concordance than the pH6.7 variant, it should be noted that all of the concordant predictions were correct with respect to the reference data.
BALB/c 3T3	Scientific basis of the test defined SOP and photo catalogue produced New statistical method applied (General linear model with a negative binomial distribution and identity as link function)	One chemical (positive control 3-methylcholanthrene) tested coded and non-coded, in 1 laboratory (for all laboratories) Same prediction (call) and dose-dependent induction of morphological transformation frequency for coded and non-coded chemical	No dedicated testing Laboratory training organised Photo catalogue produced during the training to support consistent scoring	Six chemicals (4 carcinogens, 2 non-carcinogens) tested in 3 laboratories Without considering repeat experiments, 3/6 chemicals were concordantly identified by the laboratories (1 chemical 2-acetylaminofluorene with inconclusive call in one laboratory but experiment not repeated, 2 chemicals phenanthrene and o-toluidine HCl with different calls among the laboratories) When considering repeated experiments for 2 chemicals (phenanthrene and o-toluidine HCl) 5/6 chemicals were concordantly identified by the laboratories (1 chemical 2-acetylaminofluorene with inconclusive call in one laboratory but experiment not repeated) Recommendations by the VMT to refine acceptance and assessment criteria

14.2 Extent to which study conclusions are justified by the study results alone

As a scientific piece of work the study is impressive. This is probably a "as good as it gets" set of experiments considering the difficulty of these assays (e.g. time and cost considerations).

Table 3 summarises how well the information requirements were addressed/assessed in view of the objective of the prevalidation study. From Table 3 the following is concluded:

Table 3: Extent to which information requirements were addressed and fulfilled in view of the objective of the prevalidation study

CTA variant	Protocol standardisation	Within-laboratory reproducibility	Transferability	Between-laboratory reproducibility
SHE pH 6.7	Achieved. Single reporting format would have been beneficial. Development of the photo catalogue is considered a major merit of the study.	Not sufficiently addressed Only one substance tested. In some cases this was the PC (intrinsic propensity to generate reproducible results). Study design of WLR phase was very variable.	Successfully transferred to experienced laboratories. Success of transfer was not tested empirically but can be deduced from information on BLR.	Satisfactorily demonstrated* Satisfactory for the substances tested*
SHE pH 7.0	Achieved Single reporting format would have been beneficial. Development of the photo catalogue is considered a major merit of the study.	Not sufficiently addressed Only one substance tested. In some cases this was the PC (intrinsic propensity to generate reproducible results). Study design of WLR phase was very variable.	Successfully transferred to experienced laboratories. Success of transfer was not tested empirically but can be deduced from information on BLR.	Satisfactorily demonstrated* Satisfactory for the substances tested*
BALB/c 3T3	Not finalised Assessment criteria were insufficient at outset of study. Further definition suggested by VMT and ESAC WG. These improvements need now to be assessed by testing.	Not sufficiently addressed Only one substance tested. In some cases this was the PC (intrinsic propensity to generate reproducible results). Study design of WLR phase was very variable.	Successfully transferred to experienced laboratories. Success of transfer was not tested empirically but can be deduced from information on BLR.	Promising but insufficient Further refinement of the test method required.

*) However due note should be taken that the number of substances was small (n=6), and that only one non-genotoxic carcinogen had been tested (vs. 3 genotoxic carcinogens).
PC: positive control; WLR: within-laboratory reproducibility; BLR: between-laboratory reproducibility

General observations:

(a) For the SHE pH6.7 and SHE pH7.0 CTA reports:

Not all the criteria for a prevalidation study are met. Important merits of the studies include the availability of standardised protocols and transferability of the CTAs to partly experienced laboratories. The between-laboratory reproducibility seems very promising for the selected chemicals.

(b) For the BALB/c 3T3 CTA report:

More work is required in order to satisfactorily address protocol standardisation (including definition of a positive response), within-laboratory reproducibility, transferability and between-laboratory reproducibility.

(c) Within-laboratory reproducibility:

For the three CTAs evaluated, the confidence in the within-laboratory reproducibility was not established because of the use of a single compound which in part was tested uncoded.

For the BALB/c 3T3 CTA, results obtained during the between-laboratory reproducibility assessment seem to indicate low within-laboratory reproducibility in the laboratories involved.

(d) Transferability:

The transfer of the SHE protocols was successful as judged by the acceptable between-laboratory reproducibility. The same holds true for the BALB/c 3T3 CTA when taking the repeat experiments into account.

(e) Between-laboratory reproducibility:

There was good between-laboratory reproducibility for the SHE CTA.

For the BALB/c 3T3 CTA, there were discordant results for two of the chemicals (two laboratories with a positive call, one laboratory with a negative call) when assessed according to the criteria included in the protocol. Only when the VMT intervened with more detailed testing instructions did the outlying laboratories obtain the same results as the other two.

Specific observations:

(a) Data collection:

The data on which the prevalidation study was based (OECD DRP) were adequate. The ESAC WG has no evidence suggesting obvious mistakes in data collection.

(b) Study objective and design:

The definition of the study objective is clear. The scientific rationale was described as far as our understanding of the cellular mechanisms involved in carcinogenesis allows it.

(c) Regulatory rationale:

The specific purpose of these tests was not defined by VMT. Similarly, in the OECD DRP, no specific purpose was defined although it is indicated that these cell assays can be used for identification of potential carcinogens.

(d) Study design:

- The labs involved in the prevalidation studies had variable experience with SHE and BALB/c 3T3 CTAs. Therefore we have no information about the challenges related to

transferability and reproducibility to inexperienced laboratories. Ease of transferability has thus not been assessed. This is regrettable since these tests contain a considerable element of judgement with regard to the parameters assessed (visual scoring) and information on transferability to naive laboratories would be of value. It is remarked in this context that, although there seems to be apparent robustness of these assays when considering the OECD DRP, most laboratories that produced data reported in the OECD DRP are most likely "expert" laboratories having considerable experience in conducting the test. However, one may also argue that the CTAs were transferred over time to many laboratories, which testifies to the transferability of the assay.

In summary, considering the nature of the readout (visual scoring), it is suggested that the SOPs contain a specific subsection on training and transfer to naive laboratories. Moreover, proficiency chemicals should be defined at some stage allowing the self-assessment of laboratories.

- It was questioned whether the criteria applied to select the test chemicals were appropriate in view of the study objective, i.e. assessment of reproducibility. The selection criteria applied may lead to a preponderance of strong positive/strong negative chemicals that may be more likely to give similar results in different laboratories. Therefore the chemicals selected may lead to an overestimation of reproducibility.
- For assessing the within-laboratory reproducibility only the positive control was used, partly coded and partly uncoded.
- There may be issues related to the specific decision rules based on statistical tests (assessment criteria) leading to the calls as positive and negative that need to be revisited. Statistical tests used (Fisher exact test) with the SHE assays may lead to increased false positive results. Statistical tests used with the BALB/c 3T3 assay seem to lead to some inexplicable results. The properties of the various statistical methods used need further evaluation.

(e) Test definition:

The ESAC WG appreciates the efforts to standardise the assay and subsequently the SOPs. The development of a photo catalogue supporting consistent scoring is considered a very positive result of this study because scoring was previously identified as a critical step in the conduct of the assay. While it is plausible that the photo catalogues did improve the consistency of scoring across laboratories, this is not yet entirely clear since this was not empirically tested (e.g. by including additional laboratories working without the catalogues during the between-laboratory reproducibility phase). The SOPs were considered to be acceptable provided some revisions including the ones recommended by the VMT are made.

(f) Data quality:

The ESAC WG has no evidence suggesting problems with data quality. The ESAC WG was informed that the independent statistician checked the data to see whether the acceptance criteria were met. This process worked well in this study, however, in the future perhaps clear distinction should be made between roles concerning data management and statistical analysis.

(g) Test materials:

The ESAC WG considers six chemicals as the absolute minimum for assessing reproducibility in a prevalidation study. There were concerns in the ESAC WG that the selected chemicals may produce a bias towards good reproducibility (see previous comments).

(h) Within-laboratory reproducibility:

The ESAC WG feels that within-laboratory reproducibility was not clearly established for any of the assays. Only a single substance, the positive control partly coded and partly uncoded, was tested for assessing the within-laboratory reproducibility. Moreover, the design for assessing the within-laboratory reproducibility was variable with regard to the different protocols assessed.

For the SHE pH6.7 assay, the identification as a positive result from the results with the positive control is a lenient test of within-laboratory reproducibility. For the SHE pH7.0 assay three repeats with complete dosing were performed.

In the SHE pH6.7 assay, single doses coded and non-coded were compared. In the SHE pH 7.0 assay, three dose response curves all derived using coded chemicals were compared.

In the BALB/c 3T3 assay, two dose response curves, one coded and one non-coded, were compared.

(i) Transferability:

The main finding from this transferability exercise was the importance of the scoring process. It is assumed that this process was significantly improved by the implementation of the photo catalogue. The success of the transfer was not assessed in separate experiments. Acceptable between-laboratory reproducibility however suggests that there was successful transfer of the assays – albeit to experienced laboratories.

(j) Between-laboratory reproducibility:

For both SHE assays some of the dose response curves varied appreciably. The final outcome after the implementation of the assessment criteria was considered reproducible. For the BALB/c 3T3 assay, the dose response curves seem less variable compared with those from the SHE assays, but refinement of the assessment criteria may be required.

14.3 Extent to which conclusions are plausible in the context of existing information

In case of the SHE assays, the ESAC WG considers the observed between-laboratory reproducibility plausible with respect to existing data on the assays. This is based upon a consideration of the extensive body of existing data produced with protocols which the ESAC WG analysis indicated are very similar and based upon the apparent robustness of the SHE assays (e.g. as reviewed in the OECD DRP). In case of the BALB/c 3T3 assay, the ESAC WG considers the observed between-laboratory reproducibility insufficiently supported by existing data on the assay. This is based upon a consideration of the substantial differences between the protocol variants reported e.g. in the OECD DRP (e.g. use of a modified medium, two stage-protocol, etc) and the present one (e.g. use of a new statistical method), in spite of the apparent robustness of the assay (e.g. as reviewed in the OECD DRP).

The results of the SHE assays are in good agreement with the existing data related to this assay:

For the SHE pH 6.7 assay these data include 1) the reproducibility evaluations of a similar protocol (LeBoeuf et al., 1989; Engelhardt et al., 2004) and, 2) the overall evaluation of the SHE data contained in the OECD DRP, which reported consistent results for 87.7% (57/65) of chemicals which had been tested in more than one laboratory (OECD, 2007).

For the SHE pH 7.0 assay these data include 1) the reproducibility evaluations of similar protocols as reported in the literature (Isfort et al., 1996c) and, 2) the overall evaluation of the SHE data contained in the OECD DRP, which reported consistent results for 87.7% (57/65) of chemicals which had been tested in more than one laboratory (OECD, 2007).

In addition, the data for both SHE variants add to the understanding of the predictive capacity of the CTA, which was previously addressed by the OECD DRP evaluation (OECD, 2007) of non-standardised protocols.

15. Recommendations

Note: This section should provide recommendations on the test method (e.g. further work, possible use) and their constituting elements (e.g. test system, prediction model, SOP).

15.1 General recommendations concerning the SHE assays

Although the present study succeeded in generating standardised protocols which appear reproducible, the SHE assays are at present not yet ready for regulatory use.

In any case, a revision of the protocols with the aim of incorporating the two SHE CTA protocols into one single protocol describing both pH variants (pH6.7 and pH7.0) should be considered. As a minimum, the two protocols should be harmonized as much as possible. Moreover, considering the nature of the readout (visual scoring), it is recommended that the SOPs contain a specific subsection on training and transfer of the assays to naive laboratories[13]. The definition of proficiency chemicals would support such transfer and help laboratories to assess whether they are capable of conducting the assay.

More importantly, the assays require still a complete description of their performance on the basis of a considerably larger set of chemicals including, if necessary for the envisaged purpose, pharmaceuticals and food additives. Future test substances should include substances that challenge the transferability and reproducibility (i.e. substances with discordant results between laboratories) as well as substances representing a range of possible mechanisms of action. Such performance characterisation should include information on (a) predictive capacity, (b) applicability and, more importantly, limitations of the assays, (c) reproducibility, as well as (d) ease of transferability.

When planning future steps of performance characterisation, the extent to which historical data (including earlier validation studies) can be taken into account, should be carefully considered, as these data could supplement or even substitute for a full new validation study of the SHE assays.

Moreover, the extent to which prospective testing is required to fully characterise the SHE assays will depend on (a) the intended purpose of the assays including a more precise concept concerning their possible regulatory use and (b) the information that may have – in the meantime – become available in the literature.

The following strategy is recommended in order to gain more robust information towards a complete test performance characterisation of the SHE cell assays especially for regulatory purposes:

STEP 1 – Analysis of existing information:

Any future activity towards the standardised/regulatory use of the SHE method should start with a critical analysis of the considerable body of existing testing information (either published or residing with stakeholders).

[13] The word 'naive laboratories' in the context of validation refers to laboratories that are inexperienced with regard to the use of a specific test method, i.e. they have not conducted this method or variants of it before.

It is conceivable that, after analysis of the historical datasets (e.g. with respect to chemical class, mechanism of action, carcinogenic potency) test performance can be satisfactorily described through retrospective validation and meta-analysis of data alone, without further need for prospective testing.

Importantly, such an analysis should also go back to original data and not only rely on processed data such as contained in the OECD DRP. Moreover, an evidence-based approach should be employed using a predefined search strategy for retrieving all relevant information and minimum acceptance criteria for data quality.

Should this analysis show that there are gaps in the existing data sets (e.g. with regard to chemical classes, transforming potency), STEP 2 or STEP 3 should be considered.

STEP 2 – Targeted prospective testing of gap substances:

Should the retrospective evaluation of existing information performed in STEP 1 not suffice for a satisfactory description of test performance in view of the intended purpose, a small and targeted prospective study should be conducted providing information on assay performance for those "gap substances" identified in STEP 1. The testing information generated during STEP 2 may then supplement the existing information compiled in STEP 1.

STEP 3 – Full prospective validation:

Should the information generated during STEP 1 and/or STEP2 not suffice for the intended purpose, a full prospective validation study should be conducted using the SHE CTA protocol(s) produced during this prevalidation study but taking into account the improvements of the SOPs as suggested by the ESAC.

15.2 Recommendations for improvement of the SOPs associated with the SHE assays

The following tables 4 and 5 provide an overview over the ESAC WG's recommendations concerning improvement of the SOPs of the SHE pH6.7 and SHE pH7.0 CTAs, respectively. Annex 4 provides detailed information on these recommendations including justifications why the improvements have been recommended, references to the sections of the ESAC WG report and citations of the relevant passages of the ESAC WG.

Table 4: Recommendations concerning improvements of the SOP of the SHE pH6.7 CTA:

Recommendation
1. Recommendation 1: Merger of the two SOP protocols into one single protocol describing both pH conditions (6.7 and 7.0)
2. Recommendation 2: Include mandatory testing requirement for mycoplasm
3. Recommendation 3: Better description of how to adjust target cell seeding in case of cytotoxic effects of the test item
4. Recommendation 4: Define parameters relating to cell passage and cell storage more precisely
5. Recommendation 5: Further refine the description of scoring parameters to facilitate a consistent approach
6. Recommendation 6: Include guidance on dose-range finding
7. Recommendation 7: Include guidance concerning retesting and the maximum number retesting runs before results should be considered inconclusive
8. Recommendation 8: Reconsider the recommended concentration for the Positive Control
9. Recommendation 9: Provide guidance for training and transfer
10. Recommendation 10: Consider a more objective approach to visual scoring: use of image analysis, coding of test dishes

Table 5: Recommendations concerning improvements of the SOP of the SHE pH7.0 CTA:

Recommendation
1. Recommendation 1: Merger of the two SOP protocols into one single protocol describing both pH conditions (6.7 and 7.0)
2. Recommendation 2: Include mandatory testing requirement for mycoplasm
3. Recommendation 3: Better description of how to adjust target cell seeding in case of cytotoxic effects of the test item
4. Recommendation 4: Define parameters relating to cell passage and cell storage more precisely
5. Recommendation 5: Better description of the prediction model
6. Recommendation 6: Include guidance on dose-range finding
7. Recommendation 7: Include guidance concerning retesting and the maximum number retesting runs before results should be considered inconclusive
8. Recommendation 8: Reconsider the recommended concentration for the Positive Control
9. Recommendation 9: Provide guidance for training and transfer
10. Recommendation 10: Consider a more objective approach to visual scoring: use of image analysis, coding of test dishes

15.3 General recommendations concerning the BALB/c 3T3 assay

The BALB/c 3T3 assay is at present and following this prevalidation study not yet ready for regulatory use requires further optimisation, including refinement of the acceptance and assessment criteria.

However, considering the specificities of the BALB/c 3T3 assay (e.g. use of a well-established cell line, no feeder cells needed so no irradiation facility required) compared to the SHE assays, further use of the refined protocol is encouraged to expand the data on assay reproducibility and the appropriateness of the assay assessment criteria (including statistical methodology used) for generating relevant predictions. These steps should precede a more complete test performance characterisation which may follow the same strategy as outlined for the SHE assays (see A).

Moreover, specific recommendations can be made for the BALB/c 3T3 CTA:

(a) The introduction of exogenous metabolic activation systems into the BALB/c 3T3 CTA would support the applicability of the assay to a broader range of chemicals.

(b) Considering the nature of the readout (visual scoring), it should be considered whether the SOPs should also contain a specific subsection on training and transfer. Moreover, proficiency chemicals should be defined at some stage allowing the self-assessment of laboratories.

(c) The recommended statistical method suggested for the BALB/c 3T3 CTA should be investigated more in depth to provide a better description of its properties and to allow its use in practice.

(d) The modifications of the protocol, including those suggested by the VMT, should be tested in further trials before standardised use is considered.

15.3 Recommendations for improvement of the SOPs associated with the BALB/c 3T3 assay

The following table (Table 5) provides an overview over the ESAC WG's recommendations concerning improvement of the SOPs of the BALB/c 3T3 CTA. Annex 4 provides detailed information on these recommendations including justifications why the improvements have been recommended, references to the sections of the ESAC WG report and citations of the relevant passages of the ESAC WG.

Table 5: Recommendations concerning improvements of the SOP of the BALB/c 3T3 CTA:

Recommendation
1. Recommendation 1: Include mandatory testing requirement for mycoplasm
2. Recommendation 2: Define parameters relating to cell passage more precisely
3. Recommendation 3: Further evaluate the proposed statistical method and refine the prediction model
4. Recommendation 4: Include guidance on dose-range finding
5. Recommendation 5: Include guidance concerning retesting and the maximum number retesting runs before results should be considered inconclusive
6. Recommendation 6: Define criteria for the vehicle control more precisely
7. Recommendation 7: Reconsider the recommended concentration for the Positive Control
8. Recommendation 8: Consider the use of exogenous metabolic activation system
9. Recommendation 9: Provide guidance for training and transfer
10. Recommendation 10: Consider a more objective approach to visual scoring: use of image analysis, coding of test dishes

16. References

Barrett JC and Ts'o PO (1978) Relationship between somatic mutation and neoplastic transformation. Proc. Natl. Acad. Sci. USA, 75: 3297-301.

Berwald Y and Sachs L (1965) In vitro transformation of normal cells to tumor cells by carcinogenic hydrocarbons. J. Natl. Cancer Inst., 35: 641-61.

Bryla P and Weyand EH (1992) Detection of PAH:DNA Adducts from Auto-oxidation Using 32P-Postlabeling. Cancer Letters, 65: 35-41.

Coecke S, Balls M, Bowe G, Davis J, Gstraunthaler G, Hartung T, Hay R, Merten OW, Price A, Schechtman L, Stacey G and Stokes W (2005) Guidance on good cell culture practice. A report of the second ECVAM task force on good cell culture practice. Altern. Lab. Anim., 33: 261-287.

Combes R, Balls M, Curren R, Fischbach M, Fusenig N, Kirkland D, Lasne A, Landolph J, LeBoeuf R, Marquardt H, McCormick J, Mueller L, Rivedal E, Sabbioni E, Tanaka N, Vasseur P and Yamasaki H (1999) Cell transformation assay as predictors of human carcinogenicity. Alter. Lab. Anim., 27: 745-67.

Cruciani V, Rast C, Durand MJ, Nguyen-Ba G and Vasseur P (1997) Comparative effects of clofibrate and methyl clofenapate on morphological transformation and intercellular communication of Syrian hamster embryo cells. Carcinogenesis, 18: 701-6.

DiPaolo JA, Nelson RL and Donovan PJ (1972) In vitro transformation of Syrian hamster embryo cells by diverse chemical carcinogens. Nature, 235: 278-80.

Dunkel VC, Pienta RJ, Sivak A and Traul KA (1981) Comparative neoplastic transformation responses of Balb/3T3 cells, Syrian hamster embryo cells, and Rauscher murine leukemia virus-infected Fischer 344 rat embryo cells to chemical compounds. J. Natl. Cancer Inst., 67: 1303-12.

Engelhardt G, Schwind KR and Mu•ler B (2004) The testing of chemicals in Syrian hamster embryo (SHE) cell transformation assay for assessment of carcinogenic potential. Toxicol. in vitro, 18: 213-8.

Gold LS and Zeiger E, Eds. (1997) Handbook of carcinogenicity and genotoxicity databases. CRC Press, Boca Raton (FL), USA 754 pp.

Harvey JS, Howe JR, Lynch AM and Rees RW (2005) The results of five coded compounds: genistein, metaproterenol, rotenone, p-anisidine and resorcinol tested in the pH 6.7 Syrian hamster embryo cell morphological transformation assay. Mutagenesis, 20: 51-6.

IARC (2009) Agents reviewed by the IARC Monographs Volumes 1-100A (alphabetical order). http://monographs.iarc.fr/ENG/Classification/Listagentsalphorder.pdf

Isfort RJ, Kerckaert GA and 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.

Jones CA, Callaham MF and Huberman E (1984) Enhancement of chemical-carcinogen-induced cell transformation in hamster embryo cells by 1 alpha,25-dihydroxycholecalciferol, the biologically active metabolite of vitamin D3. Carcinogenesis, 5: 1155-9.

Jones CA, Huberman E, Callaham MF, Tu A, Hallowell W, Pallota S, Sivak A, Lubet RA, Avery MD and Kouri RE (1988) An interlaboratory evaluation of the Syrian hamster embryo cell transformation assay using eighteen coded chemicals. Toxicol. In Vitro, 2: 103-116.

Kerckaert GA, Isfort RJ, Carr GJ, Aardema MJ and LeBoeuf RA (1996) A comprehensive protocol for conducting the Syrian hamster embryo cell transformation assay at pH 6.70. Mutat. Res., 356: 65-84.

LeBoeuf RA, Kerckaert KA, Poiley JA and Raineri R (1989) An interlaboratory comparison of enhanced morphological transformation of Syrian hamster embryo cells cultured under conditions of reduced bicarbonate concentration and pH. Mutat. Res., 222: 205-18.

Lu LJ (1990) Transplacental DNA damage: effects of chemical types, gestation stage, Ah Locus, and species. J Cancer Res. Clin. Oncol., 116(Suppl Pt 2): 1002.

Mauthe RJ, Gibson DT, Bunch RT and Custer L (2001) The Syrian Hamster Embryo (SHE) Cell Transformation Assay: Review of the Methods and Results. Toxicol Pathol. 29 Suppl:138-46.

Moorthy B, Chen S, Li D and Randerath K (1993) 3-Methylcholanthrene-inducible liver cytochrome(s) P450 in female Sprague-Dawley rats: possible link between P450 turnover and formation of DNA adducts and I-compounds. Carcinogenesis. 14: 879-86.

Moorthy B, Muthiah K, Fazili IS, Kondraganti SR, Wang L, Couroucli XI and Jiang W (2007) 3-Methylcholanthrene elicits DNA adduct formation in the CYP1A1 promoter region and attenuates reporter gene expression in rat H4IIE cells. Biochem. Biophys. Res. Commun., 354: 1071-7.

NTP, NTP website at http://www.ntp-server.niehs.nih.gov.

OECD (2005) Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD Series on testing and assessment Nr. 34.

OECD (2007) Detailed review paper on cell transformation assays for detection of chemical carcinogens, OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No. 31. http://appli1.oecd.org/olis/2007doc.nsf/linkto/env-jm-mono(2007)18.

Pienta RJ, Poiley JA and Lebherz WB 3rd (1977) Morphological transformation of early passage golden Syrian hamster embryo cells derived from cryopreserved primary cultures as a reliable in vitro bioassay for identifying diverse carcinogens. Int. J. Cancer, 19: 642-55.

Ridder GM, Stuard SB, Kerckaert GA, Cody DB, LeBoeuf RA, Isfort RJ (1997) Computerized image analysis of morphologically transformed and nontransformed Syrian hamster embryo (SHE) cell colonies: application to objective SHE cell transformation assay scoring. Carcinogenesis, 18: 1965-72.

Rihn BH, Bottin MC, Coulais C, Rouget R, Monhoven N, Baranowski W, Edorh A and Keith G (2000) Genotoxicity of 3-methylcholanthrene in liver of transgenic big Blue mice. Environ. Mol. Mutagen., 36: 266-73.

Rivedal E and Sanner T (1981) Metal salts as promoters of in vitro morphological transformation of hamster embryo cells initiated by benzo(a)pyrene. Cancer Res., 41: 2950-3.

Trevisan J, Angelov PP, Patel II, Najand GM, Cheung KT, Llabjani V, Pollock HM, Bruce SW, Pant K, Carmichael PL, Scott AD and Martin FL (2010) Syrian hamster embryo (SHE) assay (pH 6.7) coupled with infrared spectroscopy and chemometrics towards toxicological assessment. Analyst, 135: 3266-72.

Tsutsui T and Barrett JC (1997) Neoplastic transformation of cultured mammalian cells by estrogens and estrogenlike chemicals. Environ. Health Perspect., 105(Suppl. 3): 619-24.

Tsutsui T, Watanabe E and Barrett JC (1993) Ability of peroxisome proliferators to induce cell transformation, chromosome aberrations and peroxisome proliferation in cultured Syrian hamster embryo cells. Carcinogenesis 14: 611-18.

Tu A, Hallowell W, Pallotta S, Sivak A, Lubet RA, Curren RD, Avery MD, Jones C, Sedita BA, Huberman E, Tennant RW, Spalding J and Kouri RE (1986) An interlaboratory comparison of transformation in Syrian hamster embryo cells with model and coded chemicals. Environ. Mutagen., 8: 77-98.

Walsh MJ, Bruce SW, Pant K, Carmichael PL, Scott AD and Martin FL (2009) Discrimination of a transformation phenotype in Syrian golden hamster embryo (SHE) cells using ATR-FTIR spectroscopy. Toxicology, 258: 33-8.

Xu M, Nelson GB, Moore JE, McCoy TP, Dai J, Manderville RA, Ross JA and Miller MS (2005) Induction of Cyp1a1 and Cyp1b1 and formation of DNA adducts in C57BL/6, Balb/c, and F1 mice following in utero exposure to 3-methylcholanthrene. Toxicol. Appl. Pharmacol., 209: 28-38.

17. Annexes

Annex 1 – ESAC WG analysis of the SHE results of the prevalidation study

A) Morphological transformation frequencies of vehicle and positive controls

Comparison of the morphological transformation frequencies of vehicle and positive controls between SHE pH6.7 and pH7.0 CTAs performed in the present prevalidation study

To analyze whether the morphological transformation frequencies (MTFs) were increased in SHE pH6.7 CTA when compared with those in SHE pH7.0 CTA, the MTFs of vehicle and positive controls between the CTAs were compared (Tables 7-9).
The MTFs of vehicle control (VC) and positive control (PC) in the case of the SHE pH6.7 or SHE pH7.0 dataset of this prevalidation study are shown in Table 7 or Table 8, respectively.

Table 7. Morphological transformation frequencies (MTFs) of vehicle control (VC_1) and positive control (PC_1) in the present prevalidation study with SHE pH6.7

Laboratory	MTF (%) in the experiment with	VC_1 [*1]	PC_1 [*2]
BASF	Benzo(a)pyrene	0.37	2.47
	Anthracene	0.24	0.99
	2,4-Diaminotoluene	0.39	1.38
	3-Methylcholanthrene	0.17	2.52
	O-Toluidine HCl	0.36	2.21
	Phthalic anhydride	0.23	1.71
Harlan CCR	Benzo(a)pyrene	0.27	2.09
	Anthracene	0.06	1.58
	2,4-Diaminotoluene	0.14	2.06
	3-Methylcholanthrene	0.28	2.30
	O-Toluidine HCl	0.13	1.84
	Phthalic anhydride	0.13	1.46
BioReliance	Benzo(a)pyrene	0.47	2.29
	Anthracene	0.42	1.28
	2,4-Diaminotoluene	0.45	2.05
	3-Methylcholanthrene	0.44	1.99
	O-Toluidine HCl	0.34	1.54
	Phthalic anhydride	0.45	2.21
Mean ± S.D.		0.30 ± 0.13 [*3]	1.89 ± 0.44 [*3]

[*1] 0.2% DMSO; [*2] 5•g/ml benzo(a)pyrene; [*3] n=18

Table 8. Morphological transformation frequencies (MTFs) of vehicle control (VC_2) and positive control (PC_2) in the present prevalidation study with SHE pH7.0

Laboratory	MTF (%) in the experiment with	VC_2 [*1]	PC_2 [*2]
Uni. Metz	Anthracene	0.40	(2.85)
	2,4-Diaminotoluene	0.39	(1.82)
	3-Methylcholanthrene	0.51	(2.46)
	O-Toluidine HCl	0.49	(3.02)
	Phthalic anhydride	0.56	(1.60)
BASF	Anthracene	0.25	1.36
	2,4-Diaminotoluene	0.28	1.70
	3-Methylcholanthrene	0.25	1.36
	O-Toluidine HCl	0.26	1.43
	Phthalic anhydride	0.26	1.43
Harlan CCR	Anthracene	0.27	1.85
	2,4-Diaminotoluene	0.20	2.34
	3-Methylcholanthrene	0.32	2.04
	O-Toluidine HCl	0.35	2.29
	Phthalic anhydride	0.44	2.66
BioReliance	Anthracene	0.21	1.03
	2,4-Diaminotoluene	0.21	1.03
	3-Methylcholanthrene	0.38	1.59
	O-Toluidine HCl	0.30	1.37
	Phthalic anhydride	0.52	1.41
Mean ± S.D.		0.34 ± 0.11 [*3]	1.66 ± 0.48 [*4]

[*1] 0.2%DMSO; [*2] 5•g/ml benzo(a)pyrene except for PC_2 in Uni. Metz (1•g/ml benzo(a)pyrene) ; [*3] n=20; [*4] n=15 (The MTFs from Univ. Metz were not included because the concentration of benzo(a)pyrene used in the university was different from that used in the other laboratories.)

When compared the MTFs of VC_1 and PC_1 in SHE pH6.7 CTA with those of VC_2 and PC_2 in SHE pH7.0 CTA, there were no significant differences in the MTFs between VC_1 and VC_2 and PC_1 and PC_2 (Table 9).

Table 9. Comparison of the morphological transformation frequencies of vehicle controls ($VC_{1and\ 2}$) and positive controls ($PC_{1\ and\ 2}$) between SHE pH6.7 and SHE pH7.0 CTAs

	SHE pH6.7	SHE pH7.0		SHE pH6.7	SHE pH7.0
	VC_1	VC_2		PC_1	PC_2
Mean	0.30	0.34	Mean	1.89	1.66
Standard deviation	0.13	0.11	Standard deviation	0.44	0.48
Number of samples	18	20	Number of samples	18	15

Between VC_1 and VC_2 (P = 0.2572); Between PC_1 and PC_2 (P = 0.1697) (Unpaired t test with Welch's correction).

The results indicate that there were no significant differences in the morphological transformation frequencies of vehicle and positive controls between the SHE pH6.7 and SHE pH7.0 CTAs.

B) Predictive capacity of the test methods used in the present prevalidation study

Although the number of chemicals tested is limited, the relevance of the test methods was analyzed (Table 10).

Table 10. The relevance of the test methods used in the present prevalidation study

Performance	Test method		
	SHE pH6.7	SHE pH7.0	BALB/c 3T3
Concordance (Accuracy)	83.3% (5/6)[*1]	100% (6/6)[*2]	83.3% (5/6)[*3]
Sensitivity	100% (4/4)	100% (4/4)	100% (4/4)[*4]
Specificity	50% (1/2)	100% (2/2)[*2]	50% (1/2)[*5]
Positive predictivity	80% (4/5)	100% (4/4)	80% (4/5)
Negative predictivity	50% (1/2)	100% (2/2)[*2]	50% (1/2)

[*1] (#)shows the number of chemicals.
[*2] One of four laboratories gave a conflicting result on phthalic anhydride.
[*3] One of three laboratories gave diverging results (positive or negative) in each of three out of six chemicals.
[*4] One of three laboratories gave a diverging result on either 2-acetylaminofluorene or o-toluidine HCl.
[*5] One of three laboratories gave a positive result on phenanthrene.

C) Comparison of transformation frequencies

Comparison of the transformation frequencies induced by low or high concentrations of chemicals used as positive controls in the prevalidation study

The morphological transformation frequencies (MTFs) induced by low or high concentrations of benzo(a)pyrene in SHE pH7.0 CTA (SHE pH7.0 CTA: pp.27-29, Tables 3-6 and Figure 2C) are shown in Table 11.

Table 11. MTFs (%) induced by low or high concentrations of benzo(a)pyrene in SHE pH7.0 CTA (SHE pH7.0 CTA: pages 27-29, Tables 3-6 and Figure 2C)

	Benzo(a)pyrene	
Laboratory	1.0 or 1.25•g/ml	5•g/ml
BASF	1.87	1.57
Harlan CCR	1.45	3.01
BioReliance	1.38	2.17
Mean ± S.D.	1.57 ± 0.27 (A)	2.25 ± 0.72 (B)

Between (A) and (B) (P=0.2658 by t-test).

There were no significant differences were observed between low (1.0 or 1.25•g/ml) and high (5•g/ml) concentrations of benzo(a)pyrene in SHE pH7.0 CTA. The result suggests that a high concentration of benzo(a)pyrene used as the positive control may not hide variations in test performance.

Conversely, although there were little, if any, significant differences between the mean number of foci per dish induced by low (1•g/ml) or high (3 or 4•g/ml) concentrations of 3-methylcholanthrene in the within-laboratory reproducibility tests with BALB/c 3T3 cells, the mean number of foci per dish in BALB/c 3T3 cells treated with 3 or 4•g/ml 3-methylcholanthrene were much higher than those in cells treated with 1•g/ml 3-methylcholanthrene (Tables 12 and 13).

Table 12. Mean number of foci per dish induced by low or high concentrations of 3-methylcholanthrene in BALB/c 3T3 CTA (BALB/c 3T3 CTA, 4 Module 2: Within-laboratory reproducibility: pages 32-34, Tables 3-8)

	Concentration of 3-methylcholanthrene	
Laboratory	1•g/ml	3 or 4•g/ml
ECVAM	1.20, 1.30	4.70, 1.50
Harlan CCR	4.40, 4.60	6.70, 5.00
HRI	3.40, 3.90	11.90, 8.90
Mean ± S.D.	3.13 ± 1.52 (A)	6.45 ± 3.62 (B)

Between (A) and (B) (P=0.026 by Mann-Whitney U test; P=0.083 by t-test).

Table 13. Mean number of foci per dish induced by 4•g/ml 3-methylcholanthrene used as positive controls (BALB/c 3T3 CTA, 6 Module 4: Between-laboratory reproducibility: pp.41-67, Tables 10-31)

Laboratory	Mean number of foci per dish	Mean ± S.D.
ECVAM	20.22, 11.40 13.10, 16.10 15.60, 11.90	14.72 ± 3.30 (n=6)
Harlan CCR	11.30, 11.50 15.00, 13.60 13.70, 10.90	12.67 ± 1.66 (n=6)
HRI	16.10, 10.33 25.00, 11.25 20.70, 16.10	16.58 ± 5.58 (n=6)
Mean ± S.D.	14.66 ± 3.99 (n=18)	

The results indicate that (1) the mean number of foci per dish varied between the within- and between-laboratory reproducibility tests, and (2) treatment of BALB/c 3T3 cells with a high concentration of 3-methylcholanthrene used as the positive control may hide variations in test performance, as described in Section 3.3 (m) p.19 in this report.

Annex 2 – Analysis of SHE protocol similarity: historical vs. prevalidation

Analysis of the degree of similarity between the historical SHE cell transformation protocols and those standardised in this study

This analysis was performed by the ESAC WG in the course of its scientific review. Due to time restraints it should not be regarded as a comprehensive analysis, but rather as a study of published protocols from a selection of the more active laboratories.

A) Assay principle: Test the effect of substances on the morphology of colony forming early passage Syrian hamster embryo (SHE) cells.

The assay principle has been unaltered for more than 40 years, and the main points in the procedure are therefore also unaltered. Some changes in protocols have however been introduced in different laboratories.

B) Variation in the SHE cell transformation protocols 1965-2010:

1. Frozen cells. The assay was originally developed in the laboratories of Leo Sachs and Joe DiPaolo using freshly prepared cells from hamster embryos. Pienta introduced in 1976 the use of frozen cells which thereafter has been adopted by basically all laboratories.
2. Medium used has been DMEM with minor modifications such as alterations in glucose (high level (4.5 g/l) or low level (1 g/l)) and with or without phenol red. No reproducible effects of these alterations have been reported.
3. Medium pH has been modified by alteration of the medium bicarbonate concentration in combination with level of CO2 in the incubator.
    i. pH 7.3 was obtained using DMEM with 3.7 g/l bicarbonate and 5% CO_2.
    ii. pH 7.0 was obtained using 2-2.2 g/l bicarbonate and 10% CO_2.
    iii. pH6.7 was obtained using 0.75 g/l bicarbonate and 10% CO_2.

    pH is by far the single most effective factor in the protocols to influence morphological transformation. The effect of altered pH has in general been that more extensively transformed colonies, both in exposed and control dishes, are formed when pH is decreased. When using pH 7.3, extremely few (if any) transformed colonies were formed in unexposed dishes. When decreasing the pH to 7.0 and 6.7, the frequency of transformed colonies in unexposed dishes increases to about 0.3-0.4%. Still, this modification has been considered advantageous since a higher number of transformed colonies also were obtained in the exposed dishes.
4. Serum. 15 - 20% fetal bovine serum has been used. Most laboratories have tested for serum batches with optimum support of plating efficiency and expression of transformed morphology.
5. Syrian hamsters are out-bred animals originating from one litter of animals. Some difference in cell quality between animals has been reported, and made it relevant to test for optimal cell preparations. The advantage of having a stable source of cells (Syrian hamsters) for this cell transformation assay can not be over-estimated.

6. DMSO is the most dominantly used solvent. When necessary acetone or other solvents have been used.
7. X-irradiation of feeder cells has been performed at 4-5000R and 40-60 000 feeder cells seeded in 2 ml medium.
8. 150-500 target cells have been seeded in 2 ml medium the day after the feeder cells. The number of seeded cells has varied due to varying plating efficiency of the cells and protocols used.
9. Chemicals have been added in 4 ml complete medium the day after the target cells. The general routine has been to leave the dishes undisturbed for 7 days. Chemical exposure strategy has also been modified for specific mechanistic studies (sequential exposures, 24h exposure).
10. Fixation of colonies has been performed in 100% methanol, and the colonies thereafter stained in 10-15% Giemsa.
11. Scoring for morphological transformation. The manual scoring of individual colonies for morphological alteration involves subjectivity. Several intra- and inter-laboratorial exercises have however concluded that following training by experienced personnel, 5-10% inter-individual variance in scored colonies occur. The use of photo catalogs showing different types of normal and transformed colonies has further improved the scoring process, and contributed to the general appreciation among participants in scoring exercises that the process is less problematic than expected.

Conclusion:
Of the different alterations introduced in the SHE cell protocols, the change in medium pH is by far the factor with the largest influence on SHE cell morphological transformation. At pH 7.3, none or close to none transformed colonies were observed in unexposed dishes, while when decreasing the pH to 7.0 and 6.7, the frequency of transformed colonies increased both in unexposed and exposed dishes. Comparative data from protocols using different pH have been shown to provide reasonable concordance of the results obtained. In table 1-3 in the OECD DRP, 48 of the chemicals could be compared for response at pH 6.7 and pH •7. 40 (83%) of the chemicals gave the same response at both protocols while 8 (17%) gave different results. Of the 8 compounds giving different results, 4 were in the category "non-carcinogenic and inconclusive for carcinogenicity", 3 were positive at pH6.7 and 5 at pH •7.

Modifications other than pH have been introduced to simplify the procedure or to improve the SHE-cell plating efficiency. It can obviously not be excluded that there exist chemicals where specific alterations in the protocol may influence the final result. There are however to our knowledge no data suggesting that such alterations have influenced the final assessment of chemicals as being SHE-cell transforming or non-transforming.

There are no elements in the pre-validation protocols suggesting that the experiments carried out under the present study stand out from the experiments reported in the OECD DRP. The three present SHE cell protocols are adopted from protocols in laboratories that have been directly involved in providing significant parts of the data presented in the OECD DRP.

The analysis was based on the following references , listed in ascending chronological order.
For full bibliographic information, see section 15:

- Berwald Y, Sachs L. J Natl Cancer Inst. 1965, 35, 641-61.
- DiPaolo JA et al. Nature. 1972, 235, 278-80.

- Pienta RJ et al. Int J Cancer. 1977, 19, 642-55.
- Barrett JC, Ts'o PO. Proc Natl Acad Sci U S A. 1978, 75, 3297-301.
- Rivedal E, Sanner T. Cancer Res. 1981, 41, 2950-3.
- Dunkel VC et al. J Natl Cancer Inst. 1981, 67, 1303-12.
- Jones CA et al. Carcinogenesis. 1984, 5, 1155-9.
- Kerckaert GA et al. Mutat Res. 1996, 356, 65-84.
- Cruciani V et al. Carcinogenesis. 1997, 18, 701-6.
- Tsutsui T et al. Int J Cancer. 1997, 70, 188-93.
- Engelhardt G et al. Toxicol In Vitro. 2004, 18, 213-8.
- Harvey JS et al. Mutagenesis. 2005, 20, 51-6.
- Walsh MJ et al.Toxicology. 2009, 258, 33-8.

Annex 3 – Presentation of study results by the VMT (copied from study reports)

The Validation Management Team has summarised the results of this prevalidation study as follows (dark red font):

"SHE pH 6.7:

The VMT concluded that in keeping with the objectives of this CTA effort, the SHE pH 6.7 CTA had been prevalidated in accordance with modules 1-4 (Hartung et al., 2004). It has been demonstrated that a standardised protocol is available that should be the basis for future use. This protocol and the assay system itself are transferable between laboratories, and reproducible within- and between-laboratories.

This conclusion is substantiated by the existing body of knowledge related to this assay. In particular, by 1) the reproducibility evaluations of a similar protocol as reported in the literature (LeBoeuf et al., 1989; Engelhardt et al., 2004) and, 2) the overall evaluation of the SHE data contained in the OECD DRP, which reported consistent results for 87.7% (57/65) of chemicals which had been tested in more than one laboratory (OECD, 2007). Moreover, the VMT concluded that with the appropriate training and the use of the photo catalogue, colony scoring was not problematic despite the concerns raised in the past.

In addition, the data produced add to the understanding of the predictive capacity (module 5) of the CTA, which was previously addressed by the OECD DRP evaluation (OECD, 2007).

The VMT supports the conclusions of the OECD DRP and the generation of an OECD SHE cell transformation test guideline.

SHE pH 7.0:

The VMT concluded that in keeping with the objectives of this CTA effort, the SHE pH 7.0 CTA had been prevalidated in accordance with modules 1-4 (Hartung et al., 2004). It has been demonstrated that a standardised protocol is available that should be the basis for future use. This protocol and the assay system itself are transferable between laboratories, and reproducible within- and between-laboratories.

This conclusion is substantiated by the existing body of knowledge related to this assay. In particular, by 1) the reproducibility evaluations of similar protocols as reported in the literature (Isfort et al., 1996c) and, 2) the overall evaluation of the SHE data contained in the OECD DRP, which reported consistent results for 87.7% (57/65) of chemicals which had been tested in more than one laboratory (OECD, 2007). Moreover, the VMT concluded that with the appropriate training and the use of the photo catalogue, colony scoring was not problematic despite the concerns raised in the past.

In addition, the data produced add to the understanding of the predictive capacity (module 5) of theCTA, which was previously addressed by the OECD DRP evaluation (OECD, 2007).

The VMT supports the conclusions of the OECD DRP and the generation of an OECD SHE cell transformation test guideline.

BALB/c 3T3:

On the basis of the outcome of this prevalidation study, an improved protocol, incorporating the recommendations made by the VMT has been developed. The recommended changes to the protocol relate only to data interpretation (refinement of acceptance and assessment criteria). More experimental data are required to allow for further refinement and evaluation of the statistical method which may impact future study design (number of plates, concentrations tested, requirements for repeat studies, etc).

If the repeated experiments and the modifications to the data interpretation of the improved protocol are taken into consideration, including the importance of considering biological relevance, it can be concluded that the assay is transferable between laboratories and, reproducible within and between laboratories. Moreover, this study demonstrated that with the appropriate training and the use of the photo catalogue, the scoring of foci was not problematic despite the concerns raised in the past. It is recommended that this improved protocol be used in the future in order to confirm its utility.

Furthermore, although limited, these prevalidation data add to the fifth module i.e. predictive capacity, which has been addressed by the OECD DRP evaluation."

Annex 4 – List of detailed recommendations from the ESAC WG report regarding the SOPs of the three CTAs

The recommendations are in bold and numbered. Below each recommendation, the relevant sections of the ESAC WG report are in cited (italics) and referenced in the right column. The middle column provides a summary justification for the recommendation (blue).

A) Detailed recommendations concerning the SOP of the SHE pH6.7 CTA

Recommendation	Justification	Corresponding section of the ESAC WG report
1. Recommendation 1: Merger of the two SOP protocols into one single protocol describing both pH conditions (6.7 and 7.0)	The two SHE protocols differ only with respect to the pH used to harvest the embryonic cells and culture them subsequently. All other protocol parameters are identical or very similar (e.g. source of medium, fixatives etc). In the interest of the use of both protocols, harmonisation of the protocols of the SHE assay pH variants or merger into one single protocol would be of benefit the consistent use of this CTA.	
The use of the same template for the different SOPs would have been useful for comparison and in view of developing, in case of the SHE assays, a common protocol by merging the pH 6.7 and pH 7.0 protocols.		3.3
Implementation of minor modifications (e.g. concerning dose selection) and alignment of the two SHE protocols are required.		12.3
A revision of the protocols with the aim of incorporating the two SHE cell protocols into one single protocol describing both pH variants (pH6.7 and pH7.0) should be considered. As a minimum, the two protocols should be harmonized as much as possible.		14.4
It is noted that the SHE pH6.7 and the SHE pH7.0 use different fixatives. For the SHE pH7.0 ethanol is being used, while for the SHE 6.7 methanol was the fixative. In view of a possible harmonisation of the SHE assay SOPs, a decision to use one of the two fixatives is recommended.		3.3
2. Recommendation 2: Include mandatory testing requirement for mycoplasm	In agreement with Good Cell Culture Practice (GCCP) (Coecke et al., 2005)	
The SOPs should specify that the cell cultures should be regularly tested for mycoplasma to exclude contamination.		3.3
3. Recommendation 3: Better description of how to adjust target cell seeding in case of cytotoxic effects of the test item	A precise approach for target cell seeding is crucial to counterbalance possible loss of colonies due to cytotoxicity and, if not adjusted appropriately, may lead to an underestimation of the carcinogenicity potential. Therefore, this procedure needs to be well described.	
Adjusted target cell seeding for SHE pH 6.7 and pH 7.0 CTAs: a clear description of the procedure is needed on how to do this in a reproducible way.		3.3
4. Recommendation 4: Define parameters relating to cell passage and cell storage more precisely	It cannot be excluded that the cells will change their properties with increasing passage number or storage duration. Precision on passage parameters including maximum admissible passage number is important.	
Clearly describe the definition of passage and the meaning of "early passage" (passage number) SHE cells, e.g. SHE cells in secondary or tertiary culture.		3.3
Provide better guidance concerning the maximum duration of storage of cryopreserved SHE cells. For example, the SHE pH6.7 CTA report states in section 4.4. p.85, 2nd paragraph that "the storage period should not exceed 24 months". It should be explained why this period should not exceed 24 months.		3.3

Continued on next page

Continued from previous page

Recommendation	Justification	Corresponding section of the ESAC WG report
5. Recommendation 5: Further refine the description of scoring parameters to facilitate a consistent approach	Colony size provides additional information on cytotoxicity and complements Relative Plating Efficiency.	
Section 3.1.2 p.82: Add the criteria of normal, slightly reduced, and greatly reduced colony size and density.		3.3
6. Recommendation 6: Include guidance on dose-range finding	Dose range finding is a crucial step for defining the test dosage. Inappropriate dose spreading may lead to missing the relevant dose.	
In the context of the dose range finding, the ESAC WG recommends to include guidance for dose selection in the SHE protocols to ensure proper use of the protocols in testing laboratories. The following approach is suggested: The highest dose should be determined base on RPE (%) and the lower doses spaced out for instance with two concentrations per log (for instance: 100, 30, 10, 3, 0.3, 0.1 etc.).		8.1
7. Recommendation 7: Include guidance concerning retesting and the maximum number retesting runs before results should be considered inconclusive	Substances that give inconclusive results after one test should be retested. Retesting should however be limited to a reasonable number.	
SHE pH6.7 CTA: It is standard practice to repeat inconclusive results (e.g. o-toluidine HCl), however, the number of retesting runs should be defined in a study and, moreover, guidance should be provided in the SOP.		8.1
8. Recommendation 8: Reconsider the recommended concentration for the Positive Control (PC)	While the PC is not used for normalisation in this assay, it should also provide information on the relative sensitivity of the test system. High concentrations are however likely to yield effects. Thus, a second PC using a lower concentration or use of a 'batch control' should be considered.	
The high concentrations used for the Positive Control may hide reduced sensitivity during test performance. The concentrations should thus be reconsidered.		13
9. Recommendation 9: Provide guidance for training and transfer	The readout of the CTAs is visual scoring. A training and transfer subsection should describe an approach to transferring the test to an inexperienced laboratory.	
It should be considered whether the SOPs should provide instructions regarding the coding of test dishes to improve objectivity of the scoring process (including positive control dishes).		14.2, 14.4
Proficiency chemicals should be defined at some stage allowing the self-assessment of laboratories.		14.2, 14.4
10. Recommendation 10: Consider a more objective approach to visual scoring: use of image analysis, coding of test dishes	The readout of the CTAs is visual scoring. As with all visual assessments there is a degree of subjectivity and possible inconsistency which should be minimised to the extent possible.	
Considering the nature of the readout (visual scoring), it is suggested that the SOPs contain a specific subsection on training and transfer to naive laboratories.		13
Despite the development of the photo catalogues supporting consistent scoring, there remains a degree of subjectivity in the assessment of colonies. The ESAC WG suggests to consider automated approaches (e.g. image analysis) as possibly more objective ways to score.		13

Continued on next page

Continued from previous page

Recommendation	Justification	Corresponding section of the ESAC WG report
11. Minor recommendation	-	
Section 2.4.4 p.78, 3rd paragraph: "10-15 % aqueous Giemsa" should be "10% aqueous Giemsa". The concentration of Giemsa was defined in the SHE pH6.7 SOP on p.80, section 2.4.11 Fixing and staining of the colonies.	-	3.3

B) Detailed recommendations concerning the SOP of the SHE pH7.0 CTA

The recommendations are in bold and numbered. Below each recommendation, the relevant sections of the ESAC WG report are in cited (italics) and referenced in the right column. The middle column provides a summary justification for the recommendation (blue).

Recommendation	Justification	Corresponding section of the ESAC WG report
1. Recommendation 1: Merger of the two SOP protocols into one single protocol describing both pH conditions (6.7 and 7.0)	The two SHE protocols differ only with respect to the pH used to harvest the embryonic cells and culture them subsequently. All other protocol parameters are identical or very similar (e.g. source of medium, fixatives etc). In the interest of the use of both protocols, harmonisation of the protocols of the SHE assay pH variants or merger into one single protocol would be of benefit the consistent use of this CTA.	
The use of the same template for the different SOPs would have been useful for comparison and in view of developing, in case of the SHE assays, a common protocol by merging the pH 6.7 and pH 7.0 protocols.		3.3
Implementation of minor modifications (e.g. concerning dose selection) and alignment of the two SHE protocols are required.		12.3
A revision of the protocols with the aim of incorporating the two SHE cell protocols into one single protocol describing both pH variants (pH6.7 and pH7.0) should be considered. As a minimum, the two protocols should be harmonized as much as possible.		14.4
It is noted that the SHE pH6.7 and the SHE pH7.0 use different fixatives. For the SHE pH7.0 ethanol is being used, while for the SHE 6.7 methanol was the fixative. In view of a possible harmonisation of the SHE assay SOPs, a decision to use one of the two fixatives is recommended.		3.3
2. Recommendation 2: Include mandatory testing requirement for mycoplasm	In agreement with Good Cell Culture Practice (GCCP) (Coecke et al., 2005)	
The SOPs should specify that the cell cultures should be regularly tested for mycoplasma to exclude contamination.		3.3
3. Recommendation 3: Better description of how to adjust target cell seeding in case of cytotoxic effects of the test item	A precise approach for target cell seeding is crucial to counterbalance possible loss of colonies due to cytotoxicity and, if not adjusted appropriately, may lead to an underestimation of the carcinogenicity potential. Therefore, this procedure needs to be well described.	
Adjusted target cell seeding for SHE pH 6.7 and pH 7.0 CTAs: a clear description of the procedure is needed on how to do this in a reproducible way.		3.3
4. Recommendation 4: Define parameters relating to cell passage and cell storage more precisely	It cannot be excluded that the cells will change their properties with increasing passage number. Precision on passage parameters including maximum admissible passage number is important.	
Clearly describe the definition of passage and the meaning of "early passage" (passage number) SHE cells, e.g. SHE cells in secondary or tertiary culture.		3.3
Provide better guidance concerning the maximum duration of storage of cryopreserved SHE cells. For example, the SHE pH6.7 CTA report states in section 4.4. p.85, 2nd paragraph that "the storage period should not exceed 24 months". It should be explained why this period should not exceed 24 months.		3.3
5. Recommendation 5: Better description of the prediction model	The threshold for spontaneous morphological transformation frequency (0.6%) is an important criterion of the prediction model and should be added to the assessment criteria as has been done in the SHE pH6.7 SOP.	
The prediction model is poorly described in the SOP.		3.3

Continued on next page

Continued from previous page

Recommendation	Justification	Corresponding section of the ESAC WG report
6. Recommendation 6: Include guidance on dose-range finding	Dose range finding is a crucial step for defining the test dosage. Inappropriate dose spreading may lead to missing the relevant dose.	
In the context of the dose range finding, the ESAC WG recommends to include guidance for dose selection in the SHE protocols to ensure proper use of the protocols in testing laboratories. The following approach is suggested: The highest dose should be determined base on RPE (%) and the lower doses spaced out for instance with two concentrations per log (for instance: 100, 30, 10, 3, 0.3, 0.1 etc.).		8.1
7. Recommendation 7: Include guidance concerning retesting and the maximum number retesting runs before results should be considered inconclusive	Substances that give inconclusive results after one test should be retested. Retesting should however be limited to a reasonable number.	
The ESAC WG noted that the number of admissible retesting runs in case of unqualified tests apparently had not been defined. This, however, should ideally be done in any type of prevalidation or validation study.		4.1
8. Recommendation 8: Reconsider the recommended concentration for the Positive Control (PC)	While the PC is not used for normalisation in this assay, it should also provide information on the relative sensitivity of the test system. High concentrations are however likely to yield effects. Thus, a second PC using a lower concentration or use of a 'batch control' should be considered.	
The high concentrations used for the Positive Control may hide reduced sensitivity during test performance. The concentrations should thus be reconsidered.		13
9. Recommendation 9: Provide guidance for training and transfer	The readout of the CTAs is visual scoring. A training and transfer subsection should describe an approach to transferring the test to an inexperienced laboratory.	
It should be considered whether the SOPs should provide instructions regarding the coding of test dishes to improve objectivity of the scoring process (including positive control dishes).		14.2, 14.4
Proficiency chemicals should be defined at some stage allowing the self-assessment of laboratories.		14.2, 14.4
10. Recommendation 10: Consider a more objective approach to visual scoring: use of image analysis, coding of test dishes	The readout of the CTAs is visual scoring. As with all visual assessments there is a degree of subjectivity and possible inconsistency which should be minimised to the extent possible.	
Considering the nature of the readout (visual scoring), it is suggested that the SOPs contain a specific subsection on training and transfer to naive laboratories.		13
Despite the development of the photo catalogues supporting consistent scoring, there remains a degree of subjectivity in the assessment of colonies. The ESAC WG suggests to consider automated approaches (e.g. image analysis) as possibly more objective ways to score.		13
11. Minor recommendation		
Section 3. p.77, line 8, section 5.2.1. p.82, line 6, section 5.2.2 p.82, line 2 and section 5.3.3 p.82, line 1: "24 hours" should be "approximately 24 hours".		3.3

C) Detailed recommendations concerning the SOP of the BALB/C 3T3 CTA

The recommendations are in bold and numbered. Below each recommendation, the relevant sections of the ESAC WG report are in cited (italics) and referenced in the right column. The middle column provides a summary justification for the recommendation (blue).

Recommendation	Justification	Corresponding section of the ESAC WG report
1. Recommendation 1: Include mandatory testing requirement for mycoplasm	In agreement with Good Cell Culture Practice (GCCP) (Coecke et al., 2005)	
Status of the cells regarding mycoplasma should be clarified.		3.3
2. Recommendation 2: Define parameters relating to cell passage more precisely	Precision on passage parameters including maximum admissible passage number is important. However, it should be specified whether this number relates to the risk that cells change their properties with increasing passage number.	
Section 3.6.1 p.26, 2nd paragraph: BALB/c 3T3 cells were "...used for the CTA within 3 to 4 passages". An explanation would be helpful as to why "within 3 to 4 passages".		3.3
3. Recommendation 3: Further evaluate the proposed statistical method and refine the prediction model	The prediction model is based on the recommended statistical method. Further use of his method and further investigation of its properties should be performed to confirm the appropriateness of the assay assessment criteria. Further testing using this protocol and the associated assay assessment criteria is encouraged to gain further information on protocol performance including the statistical method.	
Refinement of the assessment requirements is required.		4.2
The criteria for an inconclusive result need strengthening.		13
The criteria for a positive / negative decision are not completely defined. It is not clear whether these should be solely based on a decision rule using a currently un-referred statistical method. The ESAC WG suggests to consider the development of a decision chart which may aid consistent and transparent decision-making.		13
Implication of comment that "experiments requiring repetition because a biological effect was observed although the statistical analysis was negative": This suggests that the design of the statistical evaluation as described in the SOPS may not be sufficient to detect effects of size that are considered biologically important. More work may need to be done on the SOPs to overcome this problem.		13
The statistical methods used requires further attention, better definition and refinement of the decision criteria based upon them. Considerations concerning the statistical methods may affect other aspects of the design of possible future prospective studies (number of plates, concentrations tested, requirements for repeat studies, etc).		14.1
There may be issues related to the specific decision rules based on statistical tests (assessment criteria) leading to the calls as positive and negative that need to be revisited. Statistical tests used with the BALB/c 3T3 assay seem to lead to some inexplicable results. The properties of the various statistical methods used need further evaluation.		14.2
The dose response curves seem less variable compared with those from the SHE assays, but refinement of the assessment criteria may be required.		14.2
The recommended statistical method suggested for the BALB/c 3T3 CTA should be investigated more in depth to provide a better description of its properties and to allow its use in practice.		14.4

Continued on next page

Continued from previous page

Recommendation	Justification	Corresponding section of the ESAC WG report
4. Recommendation 4: Include guidance on dose-range finding	Dose range finding is a crucial step for defining the test dosage. Inappropriate dose spreading may lead to missing the relevant dose.	
The doses may have to be spaced differently for the BALB/c assay since the responsive concentrations are spaced in a narrower band as compared to the SHE. This could be the reason why repetitions of BALB/c experiments with more narrow doses resulted in change from 'negative' to 'positive' calls/predictions. Based on the study data, there is therefore a need to optimize the number and spacing of doses in the BALB/c 3T3 SOP (e.g. 8 doses with what spacing?). For instance, in the BALB/c 3T3 CTA report some of the dose-responses are quite steep and could be missed if the dose spacing is not correct.		8.1
5. Recommendation 5: Include guidance concerning retesting and the maximum number retesting runs before results should be considered inconclusive	Substances that give inconclusive results after one test should be retested. Retesting should however be limited to a reasonable number.	
The ESAC WG noted that the number of admissible retesting runs in case of unqualified tests apparently had not been defined. This, however, should ideally be done in any type of prevalidation or validation study.		4.1
6. Recommendation 6: Define criteria for the vehicle control more precisely	An acceptable background level of spontaneous foci should be defined and justified. If this criterion should be based on historical data, inexperienced laboratories should be given instructions to build their own database in a reliable way.	
The maximum number of Type III foci in the entire set of vehicle control dishes should not exceed five". The specific reason for requiring a maximum of "five" foci should be explained and justified.		2.4
The acceptable number of vehicle control (VC) foci is <6. This perhaps should be justified in more detail.		2.5
7. Recommendation 7: Reconsider the recommended concentration for the Positive Control (PC)	While the PC is not used for normalisation in this assay, it should also provide information on the relative sensitivity of the test system. High concentrations are however likely to yield effects. Thus, a second PC using a lower concentration or use of a 'batch control' should be considered.	
A high dose was used for the positive control which may hide variations in test performance.		3.3
8. Recommendation 8: Consider the use of exogenous metabolic activation system	Some compounds need prior bioactivation to exert their transforming properties. The BALB/c 3T3 cells are known to have a limited metabolic activity which could restrict their use to certain compounds only. This could be mitigated by the addition of an exogenous metabolic activation system to the test system.	
The introduction of exogenous metabolic activation systems into the BALB/c 3T3 CTA would support the applicability of the assay to a broader range of chemicals.		14.4
9. Recommendation 9: Provide guidance for training and transfer	The readout of the CTAs is visual scoring. A training and transfer subsection should describe an approach to transferring the test to an inexperienced laboratory.	
It should be considered whether the SOPs should provide instructions regarding the coding of test dishes to improve objectivity of the scoring process (including positive control dishes).		13
Proficiency chemicals should be defined at some stage allowing the self-assessment of laboratories.		14.2, 14.4

Continued on next page

Continued from previous page

Recommendation	Justification	Corresponding section of the ESAC WG report
10. Recommendation 10: Consider a more objective approach to visual scoring: use of image analysis, coding of test dishes	The readout of the CTAs is visual scoring. As with all visual assessments there is a degree of subjectivity and possible inconsistency which should be minimised to the extent possible.	
Considering the nature of the readout (visual scoring), it is suggested that the SOPs contain a specific subsection on training and transfer to naïve laboratories.		14.2, 14.4
Despite the development of the photo catalogues supporting consistent scoring, there remains a degree of subjectivity in the assessment of colonies. The ESAC WG suggests to consider automated approaches (e.g. image analysis) as possibly more objective ways to score.		13
11. Minor recommendation		
p.95, Appendix 1, Culture vessels: 100x20 mm dishes should be 90x20 mm dishes.		3.3

Annex 5 – ESAC Request concerning the review of the CTA prevalidation study

ESAC Request 2010-02

ECVAM Scientific Advisory Committee

(ESAC)

ECVAM REQUEST FOR ESAC ADVICE

on an ECVAM-coordinated prevalidation study concerning the protocols of three Cell Transformation Assays (CTA) for carcinogenicity testing

INSTRUCTIONS FOR IVM/ST STAFF:

Blue text: to be filled in by the ECVAM Scientific Officer completing the draft request in collaboration with ESAC Secretariat.

Green text: to be filled in by the ESAC Secretariat.

Title page information	
Abbreviated title of ESAC request	ESAC peer review of and ESAC opinion on the ECVAM-led prevalidation study of three cell transformation assays for carcinogenicity testing: 1) SHE pH 6.7 assay 2) SHE pH 7.0 assay 3) BALB/c 3T3 assay.
ESAC REQUEST Nr.	2010-02
Filename	ESAC REQUEST_2010 02_CTA+ESAC-WG-Mandate-approved.doc
Template used for preparing request	EP 2.01
Date of finalising request	2010-09-14
Date of submitting request to ESAC	2010-10-02
Request discussed through	Plenary at ESAC33 at 2010/10/12
Opinion expected at (date)	Through written procedure in January 2011 (before OECD WNT in March)

1. TYPE OF REQUEST

Request Type	Identify request ("YES")
R1 ESAC Peer Review of a Prevalidation Study or Validation Study	YES
If R1)applies please specify further:	
• Prevalidation Study	YES The study is a complement to the recommendations of the OECD Detailed Review Paper on Cell Transformation Assays. The study addressed protocol standardisation, transferability and reproducibility (but not performance) of three protocols of cell transformation assays in view of establishing standardised protocols for future consistent use, e.g. through the development of OECD test guidelines for in vitro carcinogenicity testing.
• Prospective Validation Study	
• Retrospective Validation Study	
• Validation Study based on Performance Standards	
R2 Scientific Advice on a test method submitted to ECVAM for validation (e.g. the test method's biological relevance etc.)	
R3 Other Scientific Advice (e.g. on test methods, their use; on technical issues such as cell culturing, stem cells etc.)	

2. TITLE OF STUDY OR PROJECT FOR WHICH SCIENTIFIC ADVICE OF THE ESAC IS REQUESTED

Prevalidation of three cell transformation assays for carcinogenicity testing:

1) SHE pH 6.7 assay
2) SHE pH 7.0 assay
3) BALB/c 3T3 assay

3. BRIEF DESCRIPTION OF THE STUDY OR PROJECT

1) Background to carcinogenicity testing and available alternative methods

The potential for a compound to induce carcinogenicity is a crucial consideration when establishing hazard and risk assessment of chemicals and pharmaceuticals in humans. To date, the standard approach to assess carcinogenicity at a regulatory level is the 2-year bioassay in rodents (OECD TG 451; Ref. 1).

Several in vitro alternatives have been developed for predicting carcinogenicity. Of these, the in vitro genotoxicity tests address only one mechanism involved in carcinogenicity, the induction of genetic damage. In contrast, in vitro Cell Transformation Assays (CTAs) have been shown to involve a multistage process that closely models some stages of in vivo carcinogenesis: CTAs can detect phenotypic changes of cultured cells as a result of exposure to test materials (i.e. chemicals, products etc.). These phenotypic/morphological changes are a result of the transformation of cultured cells which involves changes in cell behaviour and proliferation control (e.g. altered cell morphology, changed colony growth patterns and anchorage –independent growth). Moreover, transformed cells can evolve to be tumorigenic when injected in a suitable host. Importantly, CTAs are to date the only optimised tests that allow the detection of both genotoxic and non-genotoxic carcinogens. CTAs have been in use for about 40 years and are currently being performed by academia, the chemical, agro-chemical, cosmetic and pharmaceutical industries. CTAs are conducted in-house as well as at contract research organisations to screen for potential carcinogenicity as well as investigate mechanisms of carcinogenicity. While CTAs are currently not used routinely for regulatory testing, they are frequently used for internal (in-house) safety assessment of chemicals, drugs, etc. and are considered to provide additional useful information to the prevailing tests that are used for assessing carcinogenic potential.

2) The OECD Detailed Review Paper as the basis for this prevalidation study

In order to systematically assess the performance of the CTAs, the Organisation for Economic Co-operation and Development (OECD) finalised in 2007 a "Detailed Review Paper on Cell Transformation Assays For Detection of Chemical Carcinogens" (OECD DRP). The OECD DRP focused on the analysis of the predictive capacity (relevance) of three CTAs and addressed also some elements of reliability: (1) the Syrian hamster embryo (SHE) assay, (2) the BALB/c 3T3 assay and (3) the C3H10T1/2 assay. A substantial body of existing and published data was evaluated (SHE n=264 chemicals; BALB/c 3T3 n=184; C3H10T1/2 n=141). The OECD DRP concluded that the performances of two of the assays, the SHE assay and BALB/c 3T3 assay, were sufficiently adequate and should be developed into formal OECD test guidelines (OECD DRP, Ref. 2). Further, the same OECD DRP recommended that although considerable data on the performance of the assays were available, a formal assessment of the assays, in particular focusing on development of a standardised transferable and reproducible protocol, would be important for preparation of such OECD test guidelines.

3) Study objectives and design

Based on the OECD DRP and several ECVAM expert meetings (Combes et al., 1999, Ref. 3), ECVAM initiated a study on the two CTAs found most relevant by the OECD DRP on the basis of the available information, the SHE and the BALB/c 3T3 assays. The study constitutes a complement to the extensive OECD DRP and its conclusions. In agreement with the conclusions of the OECD DRP, ECVAM

focused on the development and evaluation of standardised, well-documented protocols that could serve as a basis for an OECD test guideline. In summary, the study was organised and designed taking into account:

- the objective of the study to address protocol standardisation and an assessment of transferability and reproducibility of the standardised CTA protocols but not their predictive capacity (which is addressed by the OECD DRP) and
- the high costs and considerable time required to perform the assays as well as the limited funding and resources which could be made available by ECVAM.

The study addressed the three classical aspects of Prevalidation: I) protocol refinement/standardisation; II) protocol transfer and III) protocol performance (ECVAM 1995, Ref.4; OECD guidance document on validation, 2005, Ref. 5). With respect to the modular approach of validation (Hartung et al., 2004, Ref. 6), the study assessed information concerning module 1) test definition, module 2) within-laboratory reproducibility, module 3) transferability, module 4) between-laboratory reproducibility.

The study addressed three variants of CTA protocols: two SHE protocol variants (cells at pH 6.7 and at pH 7.0, respectively) and the CTA based on the BALB/c 3T3 A31 cell line. Each protocol was assessed using six chemicals. In contrast to the BALB/c 3T3 protocol which required more substantial refinement, both SHE protocols were already available in the literature and results of these have been reported in the OECD DRP. Between-laboratory reproducibility was examined in three laboratories except for the SHE 7.0 protocol, where four laboratories were involved.

4) Results and Conclusions

The Validation Management Team (VMT) concluded that, for the SHE pH 6.7 and the SHE pH 7.0 CTAs, the study had demonstrated that standardised protocols were available which could be the basis for future use. These protocols and the assay system itself have been shown to be transferable between laboratories, and are reproducible within- and between-laboratories. For the BALB/c 3T3 method, an improved protocol has been developed, which allowed obtaining reproducible results. However, further testing of the improved BALB/c 3T3 protocol is recommended (see Validation Study Reports, Ref. 7-9). Moreover, the VMT concluded that the appropriate training and the use of the photo catalogues (see Photo Catalogues, Ref. 10-12) developed during the protocol refinement phase, led to a consistent scoring of transformed colonies and foci.

Overall, these results in combination with the extensive database summarized in the OECD DRP support the utility of in vitro CTAs for the assessment of carcinogenicity potential.

References

1 OECD TG 451 on rodent long term carcinogenicity testing

2 OECD. Detailed Review Paper on Cell Transformation Assays for Detection of Chemical Carcinogens, OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No. 31 (2007).

3 Combes R., Balls M., Curren R., Fischbach M., Fusenig N., Kirkland D., Lasne A., Landolph J., LeBoeuf R., Marquardt H., McCormick J., Mueller L., Rivedal E., Sabbioni E., Tanaka N., Vasseur P. and Yamasaki H. Cell transformation assay as predictors of human carcinogenicity. Alter. Lab. Anim., 27 (1999) 745-67.

4 ECVAM Prevalidation Task Force Report 1: The role of prevalidation in the development, validation and acceptance of alternative methods. ATLA 23, 211-217 (1995)

5 OECD Series on Testing and Assessment Number 3: Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD, Paris, 2005.

6	Hartung T., Bremer S., Casati S., Coecke S., Corvi R., Fortaner S., Gribaldo L., Halder M., Hoffmann S., Roi A.J., Prieto P., Sabbioni E., Scott L., Worth A. and Zuang V. A modular approach to the ECVAM principles on test validity. Alter. Lab. Anim., 32 (2004) 467-72.
7	VMT-study report SHE pH 6.7
8	VMT-study report SHE pH 7.0
9	VMT-study report BALB/c 3T3
10	CTA SHE pH 6.7 photo catalogue
11	CTA SHE pH 7.0 photo catalogue
12	CTA BALB/c 3T3 photo catalogue

4. OBJECTIVES, QUESTIONS, TIMELINES

4.1 OBJECTIVE

Objective Why does ECVAM require advice on the current issue?	Given the background in Section 3, the opinion of the ESAC should provide expert advice to ECVAM on three studies that ECVAM conducted in view of assessing whether the three CTA protocols (SHE 6.7; SHE 7.0 and BALB/c) have been sufficiently standardised to be transferable to other laboratories and reproducible between different laboratories and may therefore be fit for future use. In providing this advice, ESAC is requested to take account of the existing information (in particular the OECD DRP) and address also the suitability of the three CTA assays/protocols in question to be used as a basis for the development of OECD test guidelines as foreseen by the OECD in the context of the OECD DRP which led to the present study.

4.2 QUESTION(S) TO BE ADDRESSED

Questions What are the questions and issues that should be addressed in view of achieving the objective of the advice?	The ESAC is requested to address the following three questions: 1) to review whether the study of the three CTAs was conducted appropriately in view of the stated purpose, i.e. of assessing whether the CTA protocols are sufficiently standardised to be transferable and reproducible. In particular the following issues should be addressed: a) Clarity of the definition of the study objective. b) Appropriateness of the study design (e.g. chemical selection, number of chemicals used, number of laboratories, acceptance criteria). c) Appropriateness of the study execution (e.g. were there pre-defined acceptance criteria, were these respected? How were exceptions / deviations handled, e.g. retesting?). d) Appropriateness of the statistical analysis as used in the protocols and for analysing reproducibility. 2) to assess whether the conclusions as presented in the Study Reports by the Validation Management Team are justified by the information generated during the study and whether they are plausible with respect to existing information and current views (e.g. literature), in particular the OECD DRP on CTAs. In particular the following issues should be addressed: a) Provide a qualitative discussion of the study results/deliverables achieved within the limits of this prevalidation study: • Clarity and completeness of the standardised protocol. • Within laboratory reproducibility • Transferability (critical issues and how they were handled)

	• Between laboratory reproducibility b) Provide a clear presentation of the conclusions presented in the study reports c) Evaluate to which extent the conclusions are justified by the study results alone d) Discuss the plausibility of the conclusion in the light of the study results AND existing historical information as available to the EWG (in particular the OECD DRP which led to this study). 3) to express its opinion with regard to the question whether the CTA protocols standardised and evaluated during the study could indeed be recommended to serve as a basis for an OECD test guideline on in vitro carcinogenicity testing. In particular the following issues should be addressed: a) Similarity of the standardised protocols with respect to the historical protocols (provide to the extent possible a direct comparison and discuss the relative importance of any difference identified). b) Other critical issues and gap analysis (what further work may be useful/required). Please provide a rationale for your proposed position.

4.3 TIMELINES

Timelines concerning this request When does ECVAM require the advice?	Timeline	Indication
	Finalised ESAC Opinion required by:	The ESAC opinion should be available latest during the second half of February 2011 (e.g. 20.1.2011). An attempt will be made to finalise the opinion by written procedure. [N.B. Following ESAC33 (12 October 2010) the entry has been revised (original text below in grey/strikethrough): ~~before February 2011, if possible (i.e. before the OECD WNT meeting)~~]
	Request to be presented to ESAC by written procedure (e.g. due to urgency) prior to the next ESAC	NO
	Request to be presented to ESAC at ESAC plenary meeting	October 2010

5. ECVAM PROPOSALS ON HOW TO ADDRESS THE REQUEST WITHIN ESAC

5.1 ECVAM PROPOSAL REGARDING REQUEST-RELATED STRUCTURES REQUIRED

Specific structures required within ESAC to address the request Does the advice require an ESAC working group, an ESAC rapporteur etc.?	Structure(s) required	Required according to ECVAM?
	S1 ESAC Rapporteur	NO
	S2 ESAC Working Group	YES
	S3 Invited Experts	NO
	Ad S3: If yes – list names and affiliations of suggested experts to be invited and specify whether these are member of the EEP	NO
	If other than above (S1-S3):	NO

5.2 DELIVERABLES AS PROPOSED BY ECVAM

Deliverables What deliverables (other than the ESAC opinion) are required for addressing the request?	Title of deliverable other than ESAC opinion	Required?
	D1 ESAC Rapporteur Report and draft opinion	NO
	D2 ESAC Peer Review Report and draft opinion	YES (ECVAM proposal)
	If other than above (D1-D2):	NO

6. LIST OF DOCUMENTS TO BE MADE AVAILABLE TO THE ESAC

Count	Description of document	Available (YES/NO)	File name
1	VMT-study report SHE pH 6.7	YES	1)ER2010-02_SHE6.7.pdf
2	VMT-study report SHE pH 7.0	YES	2)ER2010-02_SHE7.0.pdf
3	VMT-study report BALB/c 3T3	YES	3)ER2010-02_Balb.pdf
4	CTA SHE pH 6.7 photo catalogue	YES	4)ER2010-02_SHE6.7-photo.pdf
5	CTA SHE pH 7.0 photo catalogue	YES	5)ER2010-02_SHE7.0-photo.pdf
6	CTA BALB/c 3T3 photo catalogue	YES	6)ER2010-02_Balb-photo.pdf
7	OECD. Detailed Review Paper on Cell Transformation Assays for Detection of Chemical Carcinogens, OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No. 31 (2007).	YES	7)ER2010-02_OECD-DRPonCTAs.pdf
8	Combes R., Balls M., Curren R., Fischbach M., Fusenig N., Kirkland D., Lasne A., Landolph J., LeBoeuf R., Marquardt H., McCormick J., Mueller L., Rivedal E., Sabbioni E., Tanaka N., Vasseur P. and Yamasaki H. Cell transformation assay as predictors of human carcinogenicity. Alter. Lab. Anim., 27 (1999) 745-67.	YES	8)ER2010-02_ECVAM-WS-Report-on-CTAs.pdf
9	OECD TG 451 on rodent long term carcinogenicity testing	YES	9)ER2010-02_OECD-TG-451.pdf
10	ECVAM Prevalidation Task Force Report 1: The role of prevalidation in the development, validation and acceptance of alternative methods. ATLA 23, 211-217 (1995)	YES	10)ER2010-02_ECVAM-prevalidation.pdf
11	OECD Series on Testing and Assessment Number 3: Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD, Paris, 2005.	YES	11)ER2010-02_OECD-GuidanceDocument.pdf
12	Hartung T., Bremer S., Casati S., Coecke S., Corvi R., Fortaner S., Gribaldo L., Halder M., Hoffmann S., Roi A.J., Prieto P., Sabbioni E., Scott L., Worth A. and Zuang V. A modular approach to the ECVAM principles on test validity. Alter. Lab. Anim., 32 (2004) 467-72.	YES	12)ER2010-02_ECVAM-modular-approach.pdf

7. TERMS OF REFERENCE OF THE ESAC WORKING GROUP

7.1 ESTABLISHMENT OF THE ESAC WORKING GROUP

During its 33rd meeting on 12 October 2010 the ESAC plenary unanimously decided to establish an ESAC Working Group charged with the detailed scientific review of a study on three Cell Transformation (CTA) protocols.

7.2 TITLE OF THE ESAC WORKING GROUP

Full title:
"ESAC Working Group on the scientific review of 3 Cell Transformation Assay (CTA) prevalidation studies (SHE 6.7, SHE 7.0, BALB)".

Abbreviated title:
"ESAC Working Group CTA"

7.3 MANDATE OF THE ESAC WG

The EWG is requested to conduct a scientific review of the ECVAM study concerning three protocols of the Cell Transformation Assay (CTA). The review needs to address the questions put forward to ESAC by ECVAM.

The review should focus on the appropriateness of design and conduct of the study in view of the study objective and should provide an appraisal to which extent the conclusions of the Validation Management Team (VMT) are substantiated by the information generated during the study and how the information generated relates to the scientific background available.

7.4 DELIVERABLE OF THE ESAC WG

The ESAC WG is requested to deliver to the chair of the ESAC and the ESAC Secretariat a detailed ESAC Working Group Report outlining its analyses and conclusions. A reporting template has been appended (Appendix 1) intended to facilitate the drafting of the report.

The conclusions drawn in the report should be based preferably on consensus. If no consensus can be achieved, the report should clearly outline the differences in the appraisals and provide appropriate scientific justifications.

7.5 PROPOSED TIMELINES OF THE ESAC WG

The Secretariat has proposed timelines which should be agreed upon during the first Teleconference (Item 1 in the table):

Item	Proposed date/time	Action	Deliverable
1	Teleconference 5 November 2010, 14:00 CET	Kick-off teleconference to • discuss the mandate, deliverables, timelines, study background • agree on timelines and meeting dates/times (see item2) • distribute (if appropriate) work and agree on further communication (e.g. TCs if required)	• Agreed timelines • Agreed work plan and distribution
2	First ESAC WG meeting in Ispra • Option 1 (preferred): 30.11. – 2.12.2010 (3 days) • Option 2: 6.12 - 7.12.2010 (2 days) • Option 3 (least preferred): 14.12. – 16.12.2010 (3 days)	• Discussions of the relevant material and preliminary analysis and possible conclusion. • Identification of unresolved issues and disagreements • Identification of process to resolve possible disagreements • Further work distribution and communication means (e.g. TCs) • Beginning of drafting process of report	Possibly preliminary versions of • ESAC WG Report
3	Teleconference 10. January 2011, 14:00 CET	Refresher teleconference (if required) to revisit the status of the work, plan what remains to be done before the second meeting.	
4	Second (last) ESAC WG meeting in Ispra • Option 1: 12.1. – 14.1.2011 (3 days) • Option 2: • 19.1. – 21.1.2011 (3 days)	Finalisation of ESAC WG Report	Final versions of • ESAC WG Report
5	Tuesday 25.1.2011	Handover of report to ESAC chair and Secretariat	Final edited versions (ready for distribution to ESAC): • ESAC WG Report

7.6 QUESTIONS WHICH SHOULD BE ADDRESSED BY THE ESAC WG

The ESAC WG is requested to address the three questions posed to the ESAC which have been broken down further in more specific questions by the ESAC chair, the chair of the ESAC WG and the Secretariat (see section 4.2).

When preparing the final ESAC WG report to address these questions, the ESAC WG is requested to use a pre-defined reporting template. This template (see appendix 1) follows ECVAM's modular approach and addresses to which extent the standard information requirements have been addressed by the study. The template allows moreover for addressing the issues specific studies outlined in section 4.2. The Secretariat will provide guidance if necessary.

APPENDIX 1 REPORTING TEMPLATE FOR THE ESAC WG REPORT

The following suggested template follows the ECVAM modular approach and allows at the same time for the description of the analysis and conclusions concerning more specific questions. The template can be used for various types of validation studies (e.g. prospective full studies, retrospective studies, performance-based studies and prevalidation studies). Depending on the study type and the objective of the study, not all sections may be applicable. However, for reasons of consistency and to clearly identify which information requirements have not been sufficiently addressed by a specific study, this template is uniformly used for the evaluation of validation studies.

Text in red is explanatory, not intended to be part of the title.
One section is clearly not applicable to the present CTA study (identified).

1. Data collection
    - 1.1 Information / data sources used (e.g. reference data)
    - 1.2 Search strategy
    - 1.3 Selection criteria applied to the available information

2. Study objective and design
    - 2.1 Clarity of the definition of the study objective
    - 2.2 Analysis of the scientific rationale provided
    - 2.3 Analysis of the regulatory rationale provided
    - 2.4 Appropriateness of the study design (selection of test items, number of test items, number of laboratories, retesting in case of unqualified tests etc.)
    - 2.5 Appropriateness of the statistical evaluation (independence of statisticians, statistical method)

3. Test definition (Module 1)
    - 3.1 Quality and completeness of the overall test definition (test system, protocol, test acceptance criteria etc.)
    - 3.2 Quality of the background provided concerning the purpose of the test method
    - 3.3 Quality of the documentation and completeness of (a) standardised protocols (SOPs) and (b) prediction models

4. Data quality
    - 4.1 Overall quality of the evaluated data
    - 4.2 Sufficiency of the evaluated data in view of the study objective
    - 4.3 Quality of the reference data for evaluating reliability and relevance[14]

5. Test materials
    - 5.1 Sufficiency of the number of evaluated test items in view of the study objective
    - 5.2 Representativeness of the test items with respect to the applicability domain

[14] OECD guidance document Nr. 34 on validation defines relevance as follows: "Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It sis the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of accuracy (concordance) of a test method."

6. Within-laboratory reproducibility (Module 2)

6.1 Assessment of repeatability and reproducibility in the same laboratory

6.2 Conclusion on within-laboratory reproducibility as assessed by the study

7. Transferability (Module 3)

7.1 Quality of design and analysis of the transfer phase

7.2 Conclusion on transferability to a second laboratory as assessed by the study

In particular: where critical issues that may impact on transferability identified or addressed?

8. Between-laboratory reproducibility (Module 4)

8.1 Assessment of reproducibility in different laboratories

8.2 Conclusion on reproducibility as assessed by the study

9. Predictive capacity (Module 5) N.B. Predictive capacity was outside the scope of the study

9.1 Adequacy of the assessment of the predictive capacity in view of the purpose

9.2 Overall relevance (biological relevance and accuracy) of the test method in view of the purpose

10. Applicability domain (Module 6) N.B. Since this study is not a full validation study, the assessment of the applicability domain is rather limited

10.1 Appropriateness of study design to conclude on applicability domain, limitations and exclusions

10.2 Quality of the description of applicability domain, limitations, exclusions

11. Performance standards (Module 7) N.B. Not applicable to the current study.

11.1 Adequacy of the proposed Essential Test Method Components

11.2 Adequacy of the Reference Chemicals

11.3. Adequacy of the defined Accuracy Values

12. Readiness for standardised use

12.1 Assessment of the readiness for regulatory purposes

12.2. Assessment of the readiness for other uses (in house screening etc.)

12.3 Critical aspects impacting on standardised use

12.4 Gap analysis

Identify, if appropriate, gaps in the study design and/or execution that impact on the stated study objective or the conclusions drawn.

13. Other considerations

Please address any other consideration you might have in relation to the proposed approach under this section.

14. Conclusions and recommendation

14.1 Summary of the study results and conclusions

14.2 Extent to which conclusions are justified by the study results alone

14.3 Extent to which conclusions are plausible in the context of existing information

14.4 Recommendations

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
European Centre for the Validation of Alternative Methods (ECVAM)

ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(ESAC)

ESAC OPINION

Based on the ESAC Peer Review of an ECVAM-coordinated prevalidation study concerning three protocols of the Cell Transformation Assay (CTA) for in vitro carcinogenicity testing

ESAC Opinion Nr.	2011-01
Relevant ESAC request Nr.	2010-02
Date of opinion	18.2.2011

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
European Centre for the Validation of Alternative Methods (ECVAM)

TABLE OF CONTENTS

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
European Centre for the Validation of Alternative Methods (ECVAM)

Executive Summary

The potential carcinogenicity of chemicals, pharmaceuticals and food additives is a toxicity effect of great concern. To date, the standard approach to assess carcinogenicity for regulatory purposes is the two year bioassay in rodents (OECD TG 451). Several in vitro alternative methods have been developed. While in vitro genotoxicity tests address only induction of genetic damage as a mechanisms leading to carcinogenicity, in vitro Cell Transformation Assays (CTAs) have been shown to recapitulate stages of in vivo carcinogenesis. Exposure of cultured cells to carcinogenic substances in the CTA can lead to cell transformation involving changes in cell behaviour/phenotype (e.g. proliferation control, altered cell morphology, changed colony growth patterns, anchorage independent growth). Transformed cells can lead to tumour formation in vivo when injected in a suitable host, underlining the biological relevance of the CTAs for carcinogenicity testing.

Continuing previous evaluations of the potential utility of the CTAs for standardised applications including regulatory testing (Combes et al., 1997), ECVAM conducted, from 2005 to 2010, a prevalidation study on three protocol variants of the CTA. Two of the three protocols were based on cells from Syrian Hamster Embryos (the 'SHE' variant of the CTA), the "SHE pH 6.7" and "SHE pH 7.0" assays. One protocol was based on the BALB/c 3T3 cell line, the "BALB/c assay". The study addressed the three aspects of prevalidation (EC-ECVAM, 1995): protocol refinement, transfer and preliminary assessment of, within the limits of a small scale study, protocol reproducibility within and between laboratories.

Following a request from ECVAM to ESAC in October 2010 (EC-ECVAM 2010d; c.f. Annex 2) for scientific advice on this study, the ESAC set up a Working Group (ESAC WG) charged with the detailed scientific peer review of this prevalidation study.

After careful peer review of the study reports (EC-ECVAM 2010 a-c) and considering the detailed ESAC WG peer review consensus report (EC-ECVAM 2011), the ESAC concludes, in agreement with the Validation Management Team (VMT), that the reliability of the BALB/c protocol was not adequately addressed in the present study.

In contrast, in case of the SHE pH 6.7 and SHE pH7.0, the study data indicate that sufficiently standardised protocols have been produced which appear

transferable. While the data relating to the assessment of within-laboratory reproducibility were considered insufficient, the data indicated acceptable between-laboratory reproducibility for the compounds tested and when considering the type of study (i.e. prevalidation).

In view of the possible standardised use of the SHE protocols including for regulatory purposes, the development of a common SHE protocol is recommended describing both pH variants (i.e. pH6.7 and pH7.0). In a next step, test performance needs to be characterised on the basis of a larger set of chemicals covering a broad range of chemical classes and mechanisms of action. This should in particular include the evaluation of more data for non-carcinogens.

However, when planning future activities, the extent to which existing testing information on such chemicals could be used to describe SHE protocol performance for a specific purpose should be considered carefully. Such information may either be published or reside with relevant stakeholders. In the opinion of the ESAC it is conceivable and plausible, considering the extensive body of information available, that historical SHE testing data could be used to arrive at a robust characterisation of the SHE test method performance to support possible standardised use, including for regulatory purposes.

This view is based, firstly, on the apparent robustness of the SHE assays as demonstrated by an analysis of published data in the OECD DRP: the predictions compiled in this report were obtained using non-standardised protocols and showed nevertheless a high degree of concordance. Secondly, an analysis carried out by the ESAC WG (cf. Annex 2 of ESAC WG Report) indicated appreciable similarity of the historical SHE protocols and the standardised protocols in this study, supporting the possible integration of prospective with existing testing data.

Finally, the ESAC recommends, that future activities towards the possible use of these assays should start with the definition of the intended purpose which is expected to facilitate a detailed and targeted characterisation of test performance (e.g. predictive capacity, limitations) on the basis of new or existing information.

1. Mandate of the ESAC

On its meeting on 12 October 2010, the ESAC was requested by ECVAM (see Annex 2) to conduct a scientific review of an ECVAM-coordinated prevalidation study on three protocols of the Cell Transformation Assay (CTA) for carcinogenicity testing ("CTA prevalidation study"). Two of the three protocols were based on cells from Syrian Hamster Embryos (the 'SHE' variant of the CTA). The two protocols differed mainly with respect to the pH of the medium in which the cells are kept (either pH 6.7 or pH 7.0) and the protocols are hereunder referred to as "SHE pH 6.7" and "SHE pH 7.0" assays. One protocol was based on the BALB/c 3T3 cell line and is referred to hereunder as "BALB/c assay". The study addressed the three aspects of prevalidation (ECVAM 1995): protocol refinement, transfer and preliminary assessment, on the basis of a small scale study, of protocol reproducibility within and between laboratories.

1.1 General objective of the advice to be given by ESAC

Given the background made available in Section 3 of the associated request of ECVAM to ESAC (EC-ECVAM 2010d; Annex 2) and all documentation made available to the ESAC (EC-ECVAM 2010d; Annex 2), the opinion of the ESAC should provide expert advice to ECVAM on a prevalidation study that ECVAM conducted in view of assessing whether three protocols of the Cell Transformation Assay (the variants were the SHE pH6.7, the SHE pH7.0 and BALB/c protocols) have been sufficiently standardised to be transferable to other laboratories and reproducible between different laboratories and may therefore be fit for future use.

In providing this advice, ESAC is requested to take account of the existing information, in particular the OECD DRP (OECD, 2007) and address also the suitability of the three CTA assays/protocols in question to be used as a basis for the development of OECD test guidelines as foreseen by the OECD in the context of the OECD DRP which led to the present study.

1.2 Questions that should be addressed by the ESAC and its Working Group

The specific questions related to this mandate are listed in section 4.2 of Annex 2 (EC-ECVAM 2010d).

1.3 Background to the ESAC Mandate

1.3.1 Background to the study

It is important to note that this ECVAM study, performed from 2005 to 2010, was planned on the background of past ECVAM activities towards the possible use of CTAs for carcinogenicity testing (Combes et al., 1997) but also on the background of the, at the time ongoing, OECD project towards the drafting of a Detailed Review Paper (DRP) on historical CTA data (the "OECD DRP") which took place from 1997 to 2007.

Both projects therefore temporally overlapped to some extent and information on progress in both projects was mutually taken into consideration: while the ECVAM studies relied to a great extent on reference data compiled in the OECD DRP, the recommendations made in the OECD DRP took already into account possible results of the prevalidation study conducted by ECVAM at that time and aiming at the refinement of selected CTA protocols.

According to the recommendations of the OECD DRP (published in 2007 while the ECVAM study was still ongoing), the present studies (if successful) should contribute to decisions regarding the

incorporation of the CTA assays into an OECD test guideline / guidelines. However, the specific purpose of the tests within the framework of an OECD test guideline was not defined in the OECD DRP.

Since the OECD DRP was to provide detailed information on the predictive capacity of the CTAS, the ECVAM study was planned as a complement to this DRP. Consequently, the study focused on the development and evaluation (i.e. transfer/reproducibility) of standardised and well documented protocols that could serve as a possible basis for an OECD test guideline. In contrast, predictive capacity, which typically is addressed to a preliminary extent during prevalidation studies, was not defined as a study objective and there is hence no description of the assays' accuracy (sensitivity/specificity) in the study reports.

However, since test items with reference data of the rodent bioassay (and in part of IARC) were used to assess reproducibility (see figure 1 of ESAC WG report), the predictive capacity of the standardised protocols could be calculated on the basis of the test items assessed (see Annex 1 of the ESAC WG report).

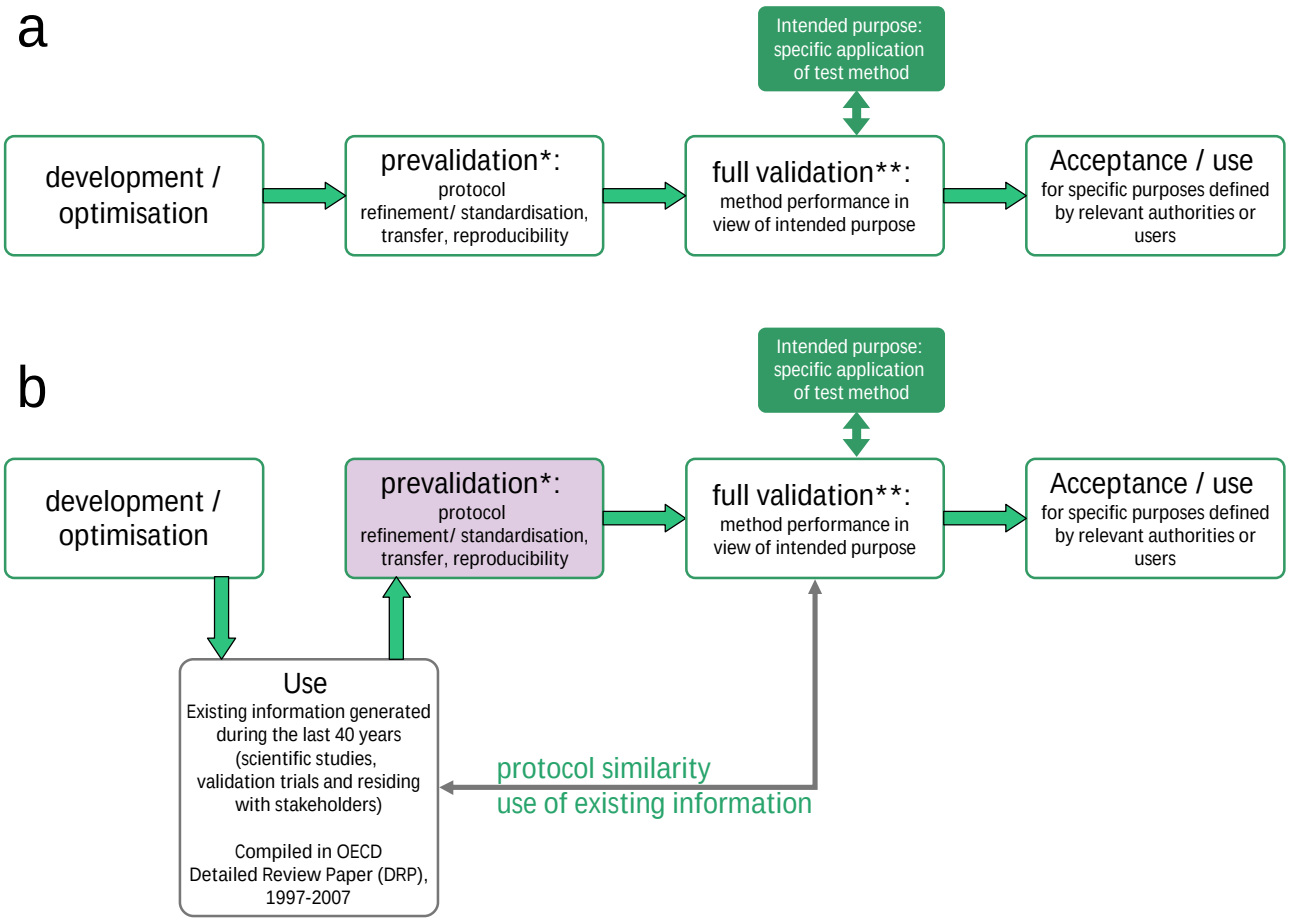
1.3.2 Specifics of this study

From the background to the study outlined in 1.3.1, it is obvious that the current prevalidation study represents a specific case (see figure 1).

Typically, a prevalidation study follows the development / optimisation phase of a new test method (figure 1a). Prevalidation studies are designed to further refine and standardise the test method's protocol (Standard Operating Procedure, SOP) and transfer the SOP from a lead laboratory to other laboratories for an assessment of between-laboratory reproducibility on the basis of a limited set of test items. This is done in order to determine whether the new method and in particular its associated protocol (SOP) is mature and robust enough to merit progressing into a costly and time-consuming full validation study which aims at a performance characterisation of the assay in view of a specific purpose.

In contrast, the present prevalidation study was planned on the background of an extensive body of data, compiled in the OECD DRP (figure 1b). The aim of the prevalidation study was, as for all prevalidation studies, the standardisation of protocols, their transfer to other laboratories and subsequent assessment of protocol reproducibility. However, in view of full test performance characterisation, the existing information and the extent to which it can be used, is an important issue and has been touched on in the recommendations (section 3.15) of this document. An important aspect in this context is the similarity of historical protocols versus protocols as generated during this study. This issue has been addressed, in a preliminary manner, by this review and is presented in more detail in Annex 2 of the ESAC WG report.

Figure 1: Validation flow. (a) Typical flow from development over prevalidation, validation to acceptance and use. (b) The situation for the CTAs. The assays have been in use for considerable time and a large body of data has been produced on the basis of the CTAs. The extent to which existing information can be used for validation and performance characterisation needs to be carefully considered.



*) Upon completion, information requirements (modular information) only partly fulfilled.
**) Upon completion, all information requirements are satisfactorily fulfilled in view of purpose

2. Summary of the ESAC Opinion

Taking into account (a) the detailed review of the ESAC WG including the WG's analysis concerning the similarity of existing protocols and those generated during the prevalidation study (EC-ECVAM 2011), (b) the information made available to ESAC by ECVAM including the Validation Study Reports (EC-ECVAM 2010 a-c), (c) the ECVAM request for ESAC advice outlining the ESAC's mandate (EC-ECVAM 2010d; c.f. see Annex 2) the ESAC has the following opinion:

(1) The reliability of the BALB/c protocol was not adequately addressed in the present study. Major concerns are, inter alia, that repeat testing was executed without blinding and also that the 'assay assessment criteria' (allowing translating the measurements into predictions on the transforming potency of substances) require further refinement as already suggested by the VMT coordinating the study. The ESAC appreciates why further refinements were made to the BALB/c protocol as a consequence of testing during the study, but believes that the final protocol having undergone these modifications should be tested in future trials.

(2) Despite shortcomings in study design and execution (see point 6), the study data indicate acceptable reliability for both the SHE pH 6.7 and SHE pH 7.0 protocols for the compounds tested and when considering the type of study (i.e. prevalidation). Reproducibility was assessed by analysing the concordance of predictions made in the participating laboratories and when comparing the predictions to in vivo carcinogenicity reference data from respected sources reviewed in the OECD Detailed Review Paper (OECD, 2007) including from IARC (IARC 2009), the National Toxicology Program (NTP database available online) and Gold and Zeiger (Gold & Zeiger 1997). Moreover, successful transfer can be concluded from the data on between-laboratory reproducibility.

(3) Although the number of test items used and consequently the chemical domain occupied is rather limited due to a shortcoming of study design (see paragraph 6) and the nature of the study (prevalidation), the ESAC is nevertheless of the opinion that it is plausible that this level of reproducibility would extend to other chemical domains as well. This notion is supported by (a) the apparent similarity of the historical versus the current protocols (c.f. analysis of protocol similarity performed by the ESAC WG, EC-ECVAM 2011) as well as (b) the apparent robustness of the CTA assays in general (see point 6 for more details). Therefore, the SHE assay protocols as standardised during this study are at least sufficiently reproducible, for those chemicals tested, to be considered for use in a regulatory setting.

(4) Despite the small number of items tested (see paragraph 6), these nevertheless reflect a certain range of the possible combinations concerning carcinogenicity and genotoxicity profiles[1]. Briefly, in case of the SHE assays, 4/6 substances tested are carcinogens (benzo(a)pyrene, 2,4-diaminotoluene, o-toluidine, 3-Methylcholantrene), while 2/6 of the substances are non-carcinogens (anthracene and phthalic anhydride). 2/4 of the carcinogenic substances are clearly genotoxic in vivo and in vitro assays (benzo(a)pyrene and 2,4-diaminotoluene), while for one the overall evidence suggests that it is a genotoxic carcinogen despite some inconclusive in vivo genotoxicity data (3-methylcholanthrene)[2]. The remaining substance (o-toluidine HCL) has equivocal data from in vivo and in vitro genotoxicity tests and could be regarded as a non-genotoxic carcinogen.

[1] A description of these mechanism-related issues relating specifically to chemical selection can be found in Vinken et al (2008).

[2] Evidence for the genotoxicity of 3-Methylcholantrene is provided for instance in: Moorthy et al., 2007; Xu et al, 2005; Rihn et al., 2000; Moorthy et al., 1993; Bryla & Wyand, 1992; for bibliographic references see section 15.

(5) While reproducibility was promising for the SHE assays, note should be taken of the fact that robust conclusions on reproducibility cannot be drawn from this prevalidation study alone. The dataset generated during this study is too small to allow sufficient characterisation of key items of test method performance. These include: reproducibility on the basis of a larger and different set of chemicals (including weakly transforming agents), predictive capacity, applicability and possible limitations.

(6) The ESAC noted that there were some shortcomings with regard to study design and execution. Although these did not critically influence the outcome of the study, they nevertheless may constitute deviations from what may be seen as good practice in validation according to ECVAM's approach and International guidance (OECD 2005). These issues include the planning and execution of the selection of reference data (see 3.4), the low number of test items (n=6), even when considering that this is a prevalidation study, the within-laboratory phase (see 3.6), the transferability phase (see 3.7) and the lack of clear provisions for retesting (e.g. number of admissible retests in case of data not fulfilling test acceptance criteria).

(7) In view of possible recommendations concerning future activities towards application of these assays for standardised testing purposes, the ESAC has the following opinion: the next step following on from this prevalidation exercise normally would be a full prospective validation study to characterise the performance of the test methods in view of standardised use including possible regulatory use. Such a study would comprise a set of test substances covering a wide range of chemical classes / possible mechanisms of action (e.g. genotoxic/non-genotoxic) and which is large enough for a statistical evaluation of predictions into two (dichotomous) classes: transforming or non-transforming agents. In line with the OECD DRP it is noted by the ESAC that also pharmaceuticals should be included in future analyses of test performance. The ESAC recommends extending this to food additives (e.g. flavours, fragrants and other food supplements).

(8) However, when planning future activities, it should be carefully considered to which extent existing information can be used. In the opinion of the ESAC it is conceivable and plausible that historical SHE testing data could be used to arrive at a robust characterisation of the SHE test method performance to support possible regulatory use. This view is based, firstly, on the apparent robustness of the SHE assays as demonstrated by an analysis of published data in the OECD DRP: the predictions compiled in this report were obtained using non-standardised protocols and showed nevertheless a high degree of concordance. Secondly, an analysis carried out by the ESAC WG (EC-ECVAM 2011: Annex 2) indicated appreciable similarity of the historical SHE protocols and the standardised protocols in this study supporting the possible integration of prospective with existing test data. About 500 coded and un-coded compounds have up to now been tested using the SHE assay by many laboratories. Careful use and reanalysis of these historical data, which include validation studies, may be able to supplement or substitute for a new full prospective validation study of the SHE assay. The ESAC WG draws attention to two publications that highlight the good predictive capacity of the SHE assays (Isfort et al, 1996; Mauthe et al, 2001).

(9) The ESAC is of the opinion, that any further activities towards assay performance characterisation should pay attention to the chemical selection and, if using existing information, to an appropriate description of the toxicity potency of substances, their physicochemical properties, chemical class and mechanism of action in order to define applicability and, in particular, possible limitations of the assays. The ESAC has two concerns with respect to the information compiled in the OECD DRP (OECD, 2007): (a) the lack of explicitness regarding the completeness of the existing data presented and whether data selection criteria based on study quality were defined and had been applied and (b) the lack of a description of the transforming potency of the chemicals analysed. Reproducibility may have been overestimated if most data are based on transforming / non-transforming agents that have generally shown unequivocal results in the past (i.e. no record of discordant results between laboratories). Thus, future activities using existing information should address these issues before

new prospective studies are planned and/or before drawing conclusions on test performance on the basis of existing data. Finally, the ESAC recommends, that future activities towards the possible use of these assays should start with the definition of the intended purpose which is expected to facilitate a detailed and targeted characterisation of test performance (e.g. predictive capacity, limitations) on the basis of new or existing information.

3. Detailed opinion of the ESAC

The following paragraphs follow largely the same structure as used in the ESAC WG report.

3.1 Data collection

3.1.1 Reference Data

All reference data used in the prevalidation study are derived from the Detailed Review Paper on Cell Transformation Assays for Detection of Chemical Carcinogens published by the OECD in 2007 (hereunder abbreviated as "OECD DRP").

The current prevalidation studies make use of data from six of the chemicals reported in the OECD DRP as reference data to assess reproducibility as measured through the concordance of predictions within and between laboratories and in reference to in vivo carcinogenicity classifications as published in the OECD DRP, which in turn – in case of these chemicals – refers to classifications by the International Agency for Research on Cancer (IARC), Gold and Zeiger (1997) and the U.S. National Toxicology Program (NTP) database.

Briefly, the predictions generated by the Cell Transformation Asssay (CTA) protocols in this study allow classification of the test chemical either as a "transforming agent" or as a "non-transforming agent" (for SHE, based on calculation of "Morphological Transformation Frequency", MTF; for BALB/c 3T3, based on measurement of the number of type III foci). Predictions (transforming / non-transforming agent) obtained in different laboratories were assessed for consistency (concordance) between the laboratories and were compared with in vivo carcinogenicity data as reported in the OECD DRP as the "reference standard".

Moreover, relevant ECVAM workshop reports and recent research papers related to (a) the scoring of observed effects, (b) the mechanistic understanding and (c) between-laboratory reproducibility are discussed in the prevalidation study reports (sections 1.4 to 1.7 pp.10-13 in all three reports), but have not been used as reference data.

For additional details, see section 1.1 of ESAC WG report

3.1.2 Search strategy to retrieve reference data associated with the test items

The ESAC working group (ESAC WG) noted that there was apparently no detailed search strategy established for identifying suitable reference data. However, taking into account that this was a small scale study which did not attempt to define the predictive capacity or the applicability domain of the three CTAs studied, but focused on protocol refinement and reproducibility, this fact was not considered relevant in this context.

For additional details, see section 1.2 of ESAC WG report

3.1.3 Selection criteria for reference data

The ESAC WG noted that there was apparently no detailed set of selection criteria established to reject/accept retrieved data. The OECD DRP was taken as reliable source although it is not clearly described in the OECD DRP how the quality of the data had been controlled.

For additional details, see section 1.3 of ESAC WG report

3.2 Study objective

3.2.1 Clarity of the study objective

The objective of the studies was considered clear and comprehensive: standardisation of CTA protocols and subsequent assessment of these protocols for reproducibility and transferability.

For additional details, see section 2.1 of ESAC WG report

3.2.2 Intended scientific rationale

The intended scientific rationale was explained as far as our current understanding of the cellular mechanisms involved in carcinogenesis (primarily in rodent cells) allows. The reported prevalidation study does not contribute to this scientific understanding, but builds upon evidence (provided primarily by the OECD DRP) that genotoxic as well as non-genotoxic carcinogens induce cell transformation in SHE and BALB/c 3T3 cells while non-carcinogenic substances do not.

For additional details, see section 2.2 of ESAC WG report

3.2.3 Regulatory rationale

The regulatory rationale remains somewhat open although it is acknowledged by the ESAC WG that even screening data and supportive data within a Weight of Evidence framework can be used for regulatory purposes and may thus constitute a "regulatory rationale". However, recommendations for a more precise definition of the regulatory usability of these tests should have been made in the reports since such use is mentioned as one of the motives for the study (see also Section 15. Recommendations).

For additional details, see section 2.3 of ESAC WG report

3.2.4 Appropriateness of study design

Overall, the study design was considered appropriate for assessing the reproducibility and transferability of the standardised protocols, despite shortcomings relating to the design/planning of (1) the test item selection (even when considering that this is a prevalidation study, there are concerns regarding the number representativeness of test chemicals with regard to chemical class and mechanism of action), (2) the within-laboratory variability phase and (3) the transferability phase.

For additional details, see section 2.4 of ESAC WG report

3.2.5 Appropriateness of statistical evaluation

The statistical evaluation of the test data generated during the study appears appropriate. However, the methods of statistical analysis used in the test method procedures (SOP) and the assay assessment criteria need a critical revision.

For additional details, see section 2.5 of ESAC WG report

3.3 Test definition

Overall the tests were adequately defined considering the objective of this study. The overall purpose of the study (development of OECD guidelines) was clear, but the specific purpose of the tests was neither defined by OECD nor by the VMT. Validation is the assessment of the satisfactory

performance of a system designed for a specific purpose. Considering this, the absence of a clearer definition of the purpose or possible use of the tests may have influenced the study design, e.g. with respect to the test chemical selection. It is therefore recommended that the intended purpose is sufficiently considered when planning a validation study including a prevalidation exercise. The SOPs were found acceptable provided some minor revisions, including the ones recommended by the VMT.

For additional details, see section 3. of ESAC WG report

3.4 Data quality

3.4.1 Quality of the evaluated data

In general, the data quality was good. Acceptance criteria are broad enough to anticipate different outcomes of the assays when applied properly. Discrepancies were explained.

3.4.2 Sufficiency of the evaluated data in view of the study objective

The data generated and evaluated did not allow for either a proper assessment of within-laboratory reproducibility or the success of the test transfer within the context of a dedicated study phase (transferability) for all three assays.

In contrast, the data produced for assessing between-laboratory reproducibility were considered sufficient for the SHE assays and may moreover be used to infer the success of transfer. The lack of appropriate testing for both modules - within-laboratory reproducibility and transferability - is, however, not considered compliant with standard practice in validation.

3.4.3 Quality of the reference data

The quality of the reference data was assumed sufficient for assessing reproducibility, being based on the OECD DRP and well-regarded sources (i.e. IARC, Gold & Zeiger and NTP database). However, it was noted that there were apparently no provision for assessing the quality of data reported in the OECD DRP and consequently in the present study.

For additional details on data quality, see section 4. of ESAC WG report

3.5 Test items

3.5.1 Sufficiency of the number of evaluated test items in view of the study objective

The number of chemicals (n=6) is judged to be sufficient, with respect to statistical requirements, to assess reproducibility (the main study objective).

However, although it is acknowledged that this is not a full validation study, the number of substances tested is low. More specifically, the number appears low to adequately cover, also for the purposes of a prevalidation study, the range of possible types of chemicals in view of the most prominent underlying mechanism of action (i.e. genotoxic / non-genotoxic) for an endpoint as complex as carcinogenicity.

Thus, the reproducibility assessment is restricted in this case to substances that belong to the same chemical classes (i.e. organic substances; inorganic compounds have not been tested) and which have the same mechanism of action as the ones tested in the prevalidation study. Considering the more advanced SHE assays, 3/6 of the substances were clear genotoxic carcinogens, only 1/6 of the substances was a possible non–genotoxic carcinogen (more details in 3.5.2).

For additional details, see section 5.1 of ESAC WG report

3.5.2 Representativeness of the test items with respect to the applicability domain

It is not the objective of this study to assess the limitations of the tests. However, based on the ESAC WG's analysis (cf. section 12.1; Annex 2) showing the apparent similarity of the historical protocols compared with the protocols from this prevalidation study, it is conceivable that historical data as reported in the OECD DRP could help describe the applicability domain of the test methods in the future.

When considering the test items of this study, it appears that mainly clear positives (transforming agents) and clear negatives (non-transforming agents) have been tested, but no "equivocal" substances (known to be able to lead to discordant results within and between laboratories) which may have challenged the reproducibility of the protocols more adequately.

For the two SHE protocols, a few materials, such as reserpine, cinnamyl anthranilate, or ethylene thiourea, which have given discordant results in previous interlaboratory studies (Tu et al. 1986; Jones et al. 1988) might have helped better define the transferability and reproducibility of these newly standardised protocols.

When considering the complexity of the endpoint (i.e. regarding decisions on "genotoxic/non-genotoxic" and "carcinogenic/non-carcinogenic") as well as considering the small number of test items (n=6), the test items covered a range of the possible combinations of (non)genotoxic and (non)carcinogenic (cf. ESAC WG report section 5.2, Box 1 'categorisation tree' and Box 2 reproducing table 35, p68 in the SHE pH6.7 assay).

Briefly, in case of the SHE assays, 4/6 substances tested are carcinogens. These are benzo(a)pyrene, 2,4-diaminotoluene, o-toluidine and 3-methylcholanthrene. 2/6 are non-carcinogens when considering reference data from the rodent bioassay (anthracene and phthalic anhydride). 1 of these non-carcinogens (anthracene) is currently not classifiable according to IARC.

Furthermore, 2/4 carcinogenic substances studied are clearly genotoxic in vivo and in vitro assays (benzo(a)pyrene and 2,4-diaminotoluene), while for one the overall evidence suggests that it is a genotoxic carcinogen despite some inconclusive in vivo genotoxicity data (3-methylcholanthrene). The remaining substance (o-toluidine HCL) has equivocal data from in vivo and in vitro genotoxicity tests and could be regarded as a non-genotoxic substance.

As the concept of non-genotoxic carcinogens has only been accepted rather recently and many substances found to be carcinogens have been tested repeatedly for genotoxicity, it is possible that such substances with equivocal genotoxicity data could be regarded as non-genotoxic carcinogens (e.g. o-toluidine).

For additional details, see section 5.2 of ESAC WG report

3.6 Within laboratory reproducibility

The ESAC WG feels that within-laboratory reproducibility was not clearly established due to inadequate study design: only one chemical was tested. Moreover, this substance was the positive control (benzo(a)pyrene for the SHE CTAs, 3-methylcholanthrene for the BALB/c 3T3 CTA). These may, due to their strong transforming potency, lead to an overestimation of reproducibility. While, within laboratory reproducibility for this single substance was, not surprisingly, high in the SHE and BALB/c assays and also between-laboratory reproducibility was good, one chemical only (in addition the PC) cannot be regarded as a sufficient dataset to conclude on this module in compliance with good validation practice.

For additional details, see section 6. of ESAC WG report

3.7 Transfer phase / Transferability

The transfer phase was adequately described and appropriately executed so allowing proper test method conduct in the other laboratories for the subsequent analysis of between laboratory reproducibility. However, how the success of the transfer was assessed and what criteria were used to judge the transfer successful was not clearly described. Moreover, ease of transferability was not assessed through the testing of test items. While this is not a prerequisite for prevalidation studies, the current study nevertheless did not fully address one of its objectives (i.e. assessment of the transferability module).

The success of the transfer programme was not demonstrated in separate experiments. All participating laboratories had some experience with CTAs. Thus, the ease of transferability to a laboratory without any CTA experience was not demonstrated. However, this is in any case not a formal requirement for a prevalidation study (OECD guidance document Nr. 34), and in this case it is noted that successful transfer may be inferred from the good between-laboratory reproducibility.

For additional details, see section 7. of ESAC WG report

3.8 Between-laboratory reproducibility

The final outcome following the implementation of the assessment criteria was considered reproducible for the SHE assays. For the BALB/c 3T3 assay, refinement of the assessment criteria is required.

Between-laboratory reproducibility was assessed through analysis of the concordance of predictions for the six test substances obtained by the involved laboratories. The predictions concerned the classification of test substances as potential transforming agents / non-transforming agents in the CTA assays. The CTA predictions were compared with the reference data associated with the test chemicals. These data are in vivo carcinogenicity predictions taken from the OECD DRP report which, for the test chemicals, are based on IARC classifications, the Gold & Zeiger and the NTP databases.

Based on the data generated and reported, the ESAC WG agrees with the VMT that the two SHE protocols yield results which are concordant between laboratories and hence reproducible for the substances tested.

In contrast, evidence supporting reproducibility of the results between laboratories for the BALB/c 3T3 protocol was considered insufficient, as suggested by the need to refine assay assessment criteria and to repeat some experiments to obtain concordant results across the laboratories.

For additional details, see section 8. of ESAC WG report

3.9 Predictive capacity

Although predictive capacity was outside the scope of the study objective, it is noteworthy that the predictions made by the SHE assays for the six chemicals were in most cases correct (6/6 corrections were correct in the SHE pH7.0, while 5/6 were correct in the SHE pH6.7). While the chemicals selected may have a bias towards reproducible results (clear negatives and strong positives), the results are nevertheless reassuring and add to the database of CTA testing data. However, based on the ESAC WG's analysis (cf. section 12.1; Annex 2) showing apparent similarity of the historical protocols with the protocols from the prevalidation study, it is conceivable that historical data as reported in the OECD DRP could help describe the predictive capacity of the SHE test methods in the future.

3.10 Applicability and limitations of the test methods

Since this study is not a full validation study, the assessment of the applicability domain is rather limited. However, based on the ESAC WG's preliminary analysis (cf. Annex 2 of ESAC WG report) showing apparent similarity of the historical protocols with the protocols from the prevalidation study, it is conceivable that historical data as reported in the OECD DRP could help describe the applicability and limitations of the SHE test methods in the future.

3.11 Performance Standards

Not applicable to this study.

3.12 Readiness for standardised use

3.12.1 Readiness for regulatory use

The data generated during this prevalidation study, when viewed on their own, are insufficient to draw conclusions on readiness for regulatory use of the SHE assay protocols, although these showed acceptable reproducibility.

However, an analysis by the ESAC WG identified considerable similarity in the historical SHE protocols and the standardised protocols in this prevalidation study. This supports the view that a substantial amount of the existing testing information from the SHE assays could be used for future considerations on their performance (e.g. predictive capacity, applicability / limitations) required to define their regulatory utility.

This view is further supported by the apparent robustness of the SHE assays as demonstrated in the OECD DRP: the data show considerable concordance with regard to the predictions made even though these older protocols may have differed to some extent and clearly no standardised test procedures had been used. These predictions are moreover relevant when compared with in vivo carcinogenicity data derived from respected sources (e.g. IARC, NTP database).

The ESAC WG, therefore, believes that future activities aimed at more precise definition of test method performance of the SHE assay and possible regulatory utility of the associated SHE protocols can be based on both, prospective testing but also on the analysis of existing historical information (e.g. a meta-analysis using defined search and data selection criteria based on study quality).

The ESAC WG notes, based upon current opinion, that no single method can provide sufficient information for an unequivocal assessment of the carcinogenicity potential of a substance to satisfy regulatory requirements fully. The SHE assays may provide information about possible genotoxic and non-genotoxic carcinogens for use in conjunction with other data (e.g. in the context of a "weight-of-evidence" approach). Some recommendations on possible approaches towards the expansion of the performance characterisation of these methods are made in section 3.15, notwithstanding the fact that the specific regulatory use needs to be defined by the relevant authorities for the purpose in mind.

The study results show that, in contrast to the SHE data, the BALB/c 3T3 protocol still requires optimisation (concerning for example the assessment criteria for the assay) and is at present neither ready to enter full validation nor consideration for regulatory use based on existing information.

For additional details, see section 12.1 of ESAC WG report

3.12.2. Assessment of the readiness for other uses

The ESAC considers the CTAs useful for testing compounds belonging to the same class of chemicals as those used in the reported prevalidation studies (screening purposes) and to generate supporting information for hazard identification and risk assessment (weight of evidence). Moreover, the CTAs will continue to be useful also for mechanistic studies of the transformation process.

For additional details, see section 12.2 of ESAC WG report

3.12.3 Critical aspects impacting on standardised use

The performance characteristics of the SHE methods need to be carefully analysed through prospective testing and/or analysis of existing information (protocol similarity supports the use of historical data) before the SHE protocols can be used in standardised applications (regulatory or non-regulatory). This analysis should include a careful examination of the chemical classes tested. Moreover, some improvement of the SHE protocols such as the development of common protocol for the two pH variants and a better description of some of the protocol steps (cell preparation) should be performed before standardised use is considered.

Concerning the BALB/c 3T3 CTA, further optimisation of the protocol is needed. These modifications, including those suggested by the VMT, should be tested in further trials before standardised use is considered.

For additional details, see section 12.3 of ESAC WG report

3.13 Other considerations

Detailed suggestions on prevalidation conduct and test method SOPs and their use have been made in the ESAC WG report (section 15.).

3.14 Conclusions on the study

As a scientific piece of work the study is impressive and succeeded in generating, in case of the SHE assays, standardised protocols including associated photo catalogues to support consistent scoring. Sufficient between-laboratory reproducibility was demonstrated for these standardised protocols. Moreover, the predictions yielded were in most cases relevant when compared to reference data (rodent bioassay and IARC class, where available).

Despite some shortcomings in study design (mainly with respect to the number of items tested, but also the design of the within-laboratory reproducibility requirement), the study succeeded with respect to its stated goals. The extent to which the various information requirements of this prevalidation study were addressed and fulfilled in view of the objective of the study is summarised in table 1.

For additional details, see section 14 of ESAC WG report

Table 1: Extent to which information requirements were addressed and fulfilled in view of the objective of the prevalidation study

	Protocol standardisation	Within-laboratory reproducibility	Transferability	Between-laboratory reproducibility
SHE pH 6.7	Achieved. Single reporting format would have been beneficial. Development of the photo catalogue is considered a major merit of the study.	Not sufficiently addressed Only one substance tested. In some cases this was the PC (intrinsic propensity to generate reproducible results). Study design of WLR phase was very variable.	Successfully transferred to experienced laboratories. Success of transfer was not tested empirically but can be deduced from information on BLR.	Satisfactorily demonstrated Satisfactory for the substances tested*
SHE pH 7.0	Achieved Single reporting format would have been beneficial. Development of the photo catalogue is considered a major merit of the study.	Not sufficiently addressed Only one substance tested. In some cases this was the PC (intrinsic propensity to generate reproducible results). Study design of WLR phase was very variable.	Successfully transferred to experienced laboratories. Success of transfer was not tested empirically but can be deduced from information on BLR.	Satisfactorily demonstrated Satisfactory for the substances tested*
BALB/c 3T3	Not finalised Assessment criteria were insufficient at outset of study. Further definition suggested by VMT and ESAC WG. These improvements need now to be assessed by testing.	Not sufficiently addressed Only one substance tested. In some cases this was the PC (intrinsic propensity to generate reproducible results). Study design of WLR phase was very variable.	Successfully transferred to experienced laboratories. Success of transfer was not tested empirically but can be deduced from information on BLR.	Promising but insufficient Further refinement of the test method required.

3.15 Recommendations for future work in view of standardised use of the test methods

As briefly outlined in section 1.3, the specific purpose of the CTAs within the framework of, for example, an OECD test guideline was neither defined in the OECD DRP nor by the Validation Management Team (VMT) when planning the current study. This has implications on the future strategy regarding test performance characterisation and possible (regulatory) use of the assays.

Test performance characteristics of a test method (e.g. applicability, limitations, predictive capacity) are to some extent dependent on the specific intended purpose of a test method. Thus, in the absence of a clear intended purpose (e.g. 'use within a test strategy for industrial chemicals to identify positives to waive confirmatory in vivo testing'), the characterisation of test performance remains difficult or on a general level which does not reflect the needs and constraints of possible applications.

Vice versa, the intended regulatory use of test method is easier to define if a precise description of test performance is available. This mutual interdependency of the test method performance characterisation and test method purpose description may, if not addressed in a forward-looking manner, hamper implementation of test methods that appear to be promising for standardised applications including regulatory testing.

In order to avoid such a situation the ESAC recommends that future activities aiming at the potential use of the SHE assays (e.g. OECD test guideline development) should commence with a definition of the intended use of the assays based on the information available (e.g. the current study and the OECD DRP). This will allow a targeted and in-depth description of test performance for the purpose in mind. A strategy towards test method characterisation has been outlined below, using existing information to the extent possible.

3.15.1. Recommendations for the SHE assays

Although the present study succeeded in generating standardised protocols which appear reproducible, the SHE assays are at present not yet ready for regulatory use.

In any case, a revision of the protocols with the aim of incorporating the two SHE cell protocols into one single protocol describing both pH variants (pH6.7 and pH7.0) should be considered. As a minimum, the two protocols should be harmonized as much as possible. Moreover, considering the nature of the readout (visual scoring), it is recommended that the SOPs contain a specific subsection on training and transfer of the assays to naïve laboratories[3]. The definition of proficiency chemicals would support such transfer and help laboratories to assess whether they are capable of conducting the assay.

More importantly, the assays require still a complete description of their performance on the basis of a considerably larger set of chemicals including, if necessary for the envisaged purpose, pharmaceuticals and food additives. Future test substances should include substances that challenge the transferability and reproducibility (i.e. substances with discordant results between laboratories) as well as substances representing a range of possible mechanisms of action. Such performance characterisation should include information on (a) predictive capacity, (b) applicability and, more importantly, limitations of the assays, (c) reproducibility, as well as (d) ease of transferability.

When planning future steps of performance characterisation, the extent to which historical data (including earlier validation studies) can be taken into account, should be carefully considered, as these data could supplement or even substitute for a full new validation study of the SHE assays.

[3] The word 'naive laboratories' in the context of validation refers to laboratories that are inexperienced with regard to the use of a specific test method, i.e. they have not conducted this method or variants of it before.

Moreover, the extent to which prospective testing is required to fully characterise the SHE assays will depend on (a) the intended purpose of the assays including a more precise concept concerning their possible regulatory use and (b) the information that may have – in the meantime – become available in the literature.

The following strategy is recommended in order to gain more robust information towards a complete test performance characterisation of the SHE cell assays especially for regulatory purposes:

- STEP 1 – Analysis of existing information:

  Any future activity towards the standardised / regulatory use of the SHE method should start with a critical analysis of the considerable body of existing testing information (either published or residing with stakeholders).

  It is conceivable that, after analysis of the historical datasets (e.g. with respect to chemical class, mechanism of action, carcinogenic potency) test performance can be satisfactorily described through retrospective validation and meta-analysis of data alone, without further need for prospective testing.

  Importantly, such an analysis should also go back to original data and not only rely on processed data such as contained in the OECD DRP. Moreover, an evidence-based approach should be employed using a predefined search strategy for retrieving all relevant information and minimum acceptance criteria for data quality.

  Should this analysis show that there are gaps in the existing data sets (e.g. with regard to chemical classes, transforming potency), STEP 2 or STEP 3 should be considered.

- STEP 2 – Targeted prospective testing of gap substances:

  Should the retrospective evaluation of existing information performed in STEP 1 not suffice for a satisfactory description of test performance in view of the intended purpose, a small and targeted prospective study should be conducted providing information on assay performance for those "gap substances" identified in STEP 1. The testing information generated during STEP 2 may then supplement the existing information compiled in STEP 1.

- STEP 3 – Full prospective validation:

  Should the information generated during STEP 1 and/or STEP2 not suffice for the intended purpose, a full prospective validation study should be conducted using the SHE protocol(s) produced during this prevalidation study but taking into account the improvements of the SOPs as suggested by the ECVAM/ESAC.

3.15.2. Recommendations for the BALB/c 3T3 assay

The BALB/c 3T3 assay is at present and following this prevalidation study not yet ready for regulatory use requires further optimisation, including refinement of the acceptance and assessment criteria.

However, considering the specificities of the BALB/c 3T3 assay (e.g. use of a well established cell line, no feeder cells needed so no irradiation facility required) compared to the SHE assays, further use of the refined protocol is encouraged to expand the data on assay reproducibility and the appropriateness of the assay assessment criteria (including statistical methodology used) for generating relevant predictions. These steps should precede a more complete test performance characterisation which may follow the same strategy as outlined for the SHE assays (see 3.15.1).

For additional details, e.g. concerning suggestions for improvement of the SOPs of the SHE and BALB/c CTAs, see section 15. of ESAC WG report

4. References

Bryla P; Weyand EH (1992) Detection of PAH:DNA Adducts from Auto-oxidation Using 32P-Postlabeling. Cancer Letters, Vol. 65, No. 1: 35-41, 25

Combes R, Balls M, Curren R, Fischbach M, Fusenig N, Kirkland D, Lasne A, Landolph J, LeBoeuf R, Marquardt H, McCormick J, Mueller L, Rivedal E, Sabbioni E, Tanaka N, Vasseur P and Yamasaki H (1999) Cell transformation assay as predictors of human carcinogenicity. Alter. Lab. Anim., 27 745-67.

EC-ECVAM (1995) ECVAM Prevalidation Task Force Report 1: The role of prevalidation in the development, validation and acceptance of alternative methods. ATLA 23, 211-217

EC-ECVAM (2010a) Report of the Validation Management Team on the ECVAM prevalidation study concerning the SHE pH6.7 Cell Transformation Assay.

EC-ECVAM (2010b) Report of the Validation Management Team on the ECVAM prevalidation study concerning the SHE pH7.0 CTA.

EC-ECVAM (2010c) Report of the Validation Management Team on the ECVAM prevalidation study concerning the BALB/c 3T3 CTA.

EC-ECVAM (2010d) ECVAM request for ESAC advice on an ECVAM-coordinated prevalidation study concerning the protocols of three Cell Transformation Assays (CTA) for in vitro carcinogenicity testing. [ESAC request 2010-02]

EC-ECVAM (2011) ESAC Working Group Consensus report on an ECVAM-coordinated prevalidation study concerning the protocols of three Cell Transformation Assays (CTA) for in vitro carcinogenicity testing.

Gold LS and Zeiger E, Eds. (1997) Handbook of carcinogenicity and genotoxicity databases. CRC Press, Boca Raton (FL), USA 754 pp.

IARC (2009) Agents reviewed by the IARC Monographs Volumes 1-100A (alphabetical order). http://monographs.iarc.fr/ENG/Classification/Listagentsalphorder.pdf

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.

Jones CA, Huberman E, Callaham MF, Tu A, Hallowell W, Pallota S, Sivak A, Lubet RA, Avery MD and Kouri RE (1988) An interlaboratory evaluation of the Syrian hamster embryo cell transformation assay using eighteen coded chemicals. Toxicol. In Vitro, 2: 103-116.

Mauthe RJ, Gibson DT, Bunch RT and Custer L (2001) The Syrian Hamster Embryo (SHE) Cell Transformation Assay: Review of the Methods and Results. Toxicol Pathol. 29 Suppl:138-46.

Moorthy B; Chen S; Li D; Randerath K (1993) 3-Methylcholanthrene-inducible liver cytochrome(s) P450 in female Sprague-Dawley rats: possible link between P450 turnover and formation of DNA adducts and I-compounds. Carcinogenesis. 14(5):879-86.

Moorthy B; Muthiah K; Fazili IS; Kondraganti SR; Wang L; Couroucli XI; Jiang W (2007) 3-Methylcholanthrene elicits DNA adduct formation in the CYP1A1 promoter region and attenuates reporter gene expression in rat H4IIE cells. Biochem Biophys Res Commun., Mar 23; 354(4):1071-7.

NTP, NTP website at http://www.ntp-server.niehs.nih.gov.

OECD (2005) Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD Series on testing and assessment Nr. 34.

OECD (2007) Detailed review paper on cell transformation assays for detection of chemical carcinogens, OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No. 31. http://appli1.oecd.org/olis/2007doc.nsf/linkto/env-jm-mono(2007)18.

Rihn BH; Bottin MC; Coulais C; Rouget R; Monhoven N; Baranowski W; Edorh A; Keith G (2000) Genotoxicity of 3-methylcholanthrene in liver of transgenic big Blue mice. Environ Mol Mutagen.; 36(4):266-73.

Lu LJ (1990) Transplacental DNA damage: effects of chemical types, gestation stage, Ah Locus, and species. J Cancer Res Clin Oncol;116(Suppl Pt 2):1002

Tu A, Hallowell W, Pallotta S, Sivak A, Lubet RA, Curren RD, Avery MD, Jones C, Sedita BA, Huberman E, Tennant RW, Spalding J and Kouri RE (1986) An interlaboratory comparison of transformation in Syrian hamster embryo cells with model and coded chemicals. Environ. Mutagen., 8: 77-98.

Vinken M, Doktorova T, Ellinger-Ziegelbauer H, Ahr HJ, Lock E, Carmichael P, Roggen E, van Delft J, Kleinjans J, Castell J, Bort R, Donato T, Ryan M, Corvi R, Keun H, Ebbels T, Athersuch T, Sansone SA, Rocca-Serra P, Stierum R, Jennings P, Pfaller W, Gmuender H, Vanhaecke T and Rogiers V (2008) The carcinoGENOMICS project: critical selection of model compounds for the development of omics-based in vitro carcinogenicity screening assays, Mutat. Res., 659: 202-10.

Xu M; Nelson GB; Moore JE; McCoy TP; Dai J; Manderville RA; Ross JA; Miller MS (2005) Induction of Cyp1a1 and Cyp1b1 and formation of DNA adducts in C57BL/6, Balb/c, and F1 mice following in utero exposure to 3-methylcholanthrene. Toxicol Appl Pharmacol. 209(1):28-38

Annex 1 ESAC and ESAC Working Group charged with the scientific review

ECVAM Scientific Advisory Committee
Dr. Nathalie ALÉPÉE
Dr. David BASKETTER
Dr. Neil CARMICHAEL
Prof. Jacques R. CHRETIEN
Prof. Lucio G. COSTA
Dr. Rodger CURREN
Prof. A. Wallace HAYES
Prof. Coenraad HENDRIKSEN
Dr. Dagmar JÍROVÁ
Prof. Walter PFALLER
Dr. Erwin ROGGEN
Prof. Vera ROGIERS
Dr. Andrea SEILER
Prof. Kristin SCHIRMER
Prof. Ruud A. WOUTERSEN

ESAC Working Group CTA
Dr. Erwin ROGGEN (ESAC member, Chair of the WG)
Dr. Rodger CURREN (ESAC member)
Dr. David LOVELL (Invited expert, EEP)
Dr. Edgar RIVEDAL (Invited expert, EEP)
Dr. Takeki TSUTSUI (Invited expert following an ICATM proposal from JaCVAM; EEP)

ESAC Secretariat
Dr. Claudius GRIESINGER
Dr. Pascal PHRAKONKHAM (specific support)

Annex 2 ECVAM request to ESAC for scientific advice concerning the ECVAM-coordinated CTA prevalidation study

ESAC Request 2010-02

ECVAM Scientific Advisory Committee

(ESAC)

ECVAM REQUEST FOR ESAC ADVICE

on an ECVAM-coordinated prevalidation study concerning the protocols of three Cell Transformation Assays (CTA) for carcinogenicity testing

Title page information	
Abbreviated title of ESAC request	ESAC peer review of and ESAC opinion on the ECVAM-led prevalidation study of three cell transformation assays for carcinogenicity testing: 1) SHE pH 6.7 assay 2) SHE pH 7.0 assay 3) Balb/c 3T3 assay.
ESAC REQUEST Nr.	2010-02
Filename	ESAC REQUEST_2010 02_CTA+ESAC-WG-Mandate-approved.doc
Template used for preparing request	EP 2.01
Date of finalising request	2010-09-14
Date of submitting request to ESAC	2010-10-02
Request discussed through	Plenary at ESAC33 at 2010/10/12
Opinion expected at (date)	Through written procedure in January 2011 (before OECD WNT in March)

TABLE OF CONTENTS

1. TYPE OF REQUEST

Request Type		Identify request ("YES")
R1 ESAC Peer Review of a Prevalidation Study or Validation Study		YES
If R1)applies please specify further:		
• Prevalidation Study	YES The study is a complement to the recommendations of the OECD Detailed Review Paper on Cell Transformation Assays. The study addressed protocol standardisation, transferability and reproducibility (but not performance) of three protocols of cell transformation assays in view of establishing standardised protocols for future consistent use, e.g. through the development of OECD test guidelines for in vitro carcinogenicity testing.	
• Prospective Validation Study		
• Retrospective Validation Study		
• Validation Study based on Performance Standards		
R2 Scientific Advice on a test method submitted to ECVAM for validation (e.g. the test method's biological relevance etc.)		
R3 Other Scientific Advice (e.g. on test methods, their use; on technical issues such as cell culturing, stem cells etc.)		

2. TITLE OF STUDY OR PROJECT FOR WHICH SCIENTIFIC ADVICE OF THE ESAC IS REQUESTED

Prevalidation of three cell transformation assays for carcinogenicity testing:

1) SHE pH 6.7 assay
2) SHE pH 7.0 assay
3) Balb/c 3T3 assay

3. BRIEF DESCRIPTION OF THE STUDY OR PROJECT

1) Background to carcinogenicity testing and available alternative methods

The potential for a compound to induce carcinogenicity is a crucial consideration when establishing hazard and risk assessment of chemicals and pharmaceuticals in humans. To date, the standard approach to assess carcinogenicity at a regulatory level is the 2-year bioassay in rodents (OECD TG 451; Ref. 1).

Several in vitro alternatives have been developed for predicting carcinogenicity. Of these, the in vitro genotoxicity tests address only one mechanism involved in carcinogenicity, the induction of genetic damage. In contrast, in vitro Cell Transformation Assays (CTAs) have been shown to involve a multistage process that closely models some stages of in vivo carcinogenesis: CTAs can detect phenotypic changes of cultured cells as a result of exposure to test materials (i.e. chemicals, products etc.). These phenotypic/morphological changes are a result of the transformation of cultured cells which involves changes in cell behaviour and proliferation control (e.g. altered cell morphology, changed colony growth patterns and anchorage –independent growth). Moreover, transformed cells can evolve to be tumorigenic when injected in a suitable host. Importantly, CTAs are to date the only optimised tests that allow the detection of both genotoxic and non-genotoxic carcinogens. CTAs have been in use for about 40 years and are currently being performed by academia, the chemical, agro-chemical, cosmetic and pharmaceutical industries. CTAs are conducted in-house as well as at contract research organisations to screen for potential carcinogenicity as well as investigate mechanisms of carcinogenicity. While CTAs are currently not used routinely for regulatory testing, they are frequently used for internal (in-house) safety assessment of chemicals, drugs, etc. and are considered to provide additional useful information to the prevailing tests that are used for assessing carcinogenic potential.

2) The OECD Detailed Review Paper as the basis for this prevalidation study

In order to systematically assess the performance of the CTAs, the Organisation for Economic Co-operation and Development (OECD) finalised in 2007 a "Detailed Review Paper on Cell Transformation Assays For Detection of Chemical Carcinogens" (OECD DRP). The OECD DRP focused on the analysis of the predictive capacity (relevance) of three CTAs and addressed also some elements of reliability: (1) the Syrian hamster embryo (SHE) assay, (2) the BALB/c 3T3 assay and (3) the C3H10T1/2 assay. A substantial body of existing and published data was evaluated (SHE n=264 chemicals; BALB/c 3T3 n=184; C3H10T1/2 n=141). The OECD DRP concluded that the performances of two of the assays, the SHE assay and Balb/c 3T3 assay, were sufficiently adequate and should be developed into formal OECD test guidelines (OECD DRP, Ref. 2). Further, the same OECD DRP recommended that although considerable data on the performance of the assays were available, a formal assessment of the assays, in particular focusing on development of a standardised transferable and reproducible protocol, would be important for preparation of such OECD test guidelines.

3) Study objectives and design

Based on the OECD DRP and several ECVAM expert meetings (Combes et al., 1999, Ref. 3), ECVAM initiated a study on the two CTAs found most relevant by the OECD DRP on the basis of the available information, the SHE and the BALB/c 3T3 assays. The study constitutes a complement to the extensive OECD DRP and its conclusions. In agreement with the conclusions of the OECD DRP, ECVAM focused on the development and evaluation of standardised, well-documented protocols that could serve as a basis for an OECD test guideline. In summary, the study was organised and designed taking into account:

- the objective of the study to address protocol standardisation and an assessment of transferability and reproducibility of the standardised CTA protocols but not their predictive capacity (which is addressed by the OECD DRP) and
- the high costs and considerable time required to perform the assays as well as the limited funding and resources which could be made available by ECVAM.

The study addressed the three classical aspects of Prevalidation: I) protocol refinement/standardisation; II) protocol transfer and III) protocol performance (ECVAM 1995, Ref.4; OECD guidance document on validation, 2005, Ref. 5). With respect to the modular approach of validation (Hartung et al., 2004, Ref. 6), the study assessed information concerning module 1) test definition, module 2) within-laboratory reproducibility, module 3) transferability, module 4) between-laboratory reproducibility.

The study addressed three variants of CTA protocols: two SHE protocol variants (cells at pH 6.7 and at pH 7.0, respectively) and the CTA based on the Balb/c 3T3 A31 cell line. Each protocol was assessed using six chemicals. In contrast to the Balb/c 3T3 protocol which required more substantial refinement, both SHE protocols were already available in the literature and results of these have been reported in the OECD DRP. Between-laboratory reproducibility was examined in three laboratories except for the SHE 7.0 protocol, where four laboratories were involved.

4) Results and Conclusions

The Validation Management Team (VMT) concluded that, for the SHE pH 6.7 and the SHE pH 7.0 CTAs, the study had demonstrated that standardised protocols were available which could be the basis for future use. These protocols and the assay system itself have been shown to be transferable between laboratories, and are reproducible within- and between-laboratories. For the Balb/c 3T3 method, an improved protocol has been developed, which allowed obtaining reproducible results. However, further testing of the improved Balb/c 3T3 protocol is recommended (see Validation Study Reports, Ref. 7-9). Moreover, the VMT concluded that the appropriate training and the use of the photo catalogues (see Photo Catalogues, Ref. 10-12) developed during the protocol refinement phase, led to a consistent scoring of transformed colonies and foci.

Overall, these results in combination with the extensive database summarized in the OECD DRP support the utility of in vitro CTAs for the assessment of carcinogenicity potential.

References

1 OECD TG 451 on rodent long term carcinogenicity testing

2 OECD. Detailed Review Paper on Cell Transformation Assays for Detection of Chemical Carcinogens, OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No. 31 (2007).

3 Combes R., Balls M., Curren R., Fischbach M., Fusenig N., Kirkland D., Lasne A., Landolph J., LeBoeuf R., Marquardt H., McCormick J., Mueller L., Rivedal E., Sabbioni E., Tanaka N., Vasseur P. and Yamasaki H. Cell transformation assay as predictors of human carcinogenicity. Alter. Lab. Anim., 27 (1999) 745-67.

4 ECVAM Prevalidation Task Force Report 1: The role of prevalidation in the development, validation and acceptance of alternative methods. ATLA 23, 211-217 (1995)

5 OECD Series on Testing and Assessment Number 3: Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD, Paris, 2005.

6 Hartung T., Bremer S., Casati S., Coecke S., Corvi R., Fortaner S., Gribaldo L., Halder M., Hoffmann S., Roi A.J., Prieto P., Sabbioni E., Scott L., Worth A. and Zuang V. A modular approach to the ECVAM principles on test validity. Alter. Lab. Anim., 32 (2004) 467-72.

7	VMT-study report SHE pH 6.7
8	VMT-study report SHE pH 7.0
9	VMT-study report Balb/c 3T3
10	CTA SHE pH 6.7 photo catalogue
11	CTA SHE pH 7.0 photo catalogue
12	CTA Balb/c 3T3 photo catalogue

4. OBJECTIVES, QUESTIONS, TIMELINES

4.1 OBJECTIVE

Objective Why does ECVAM require advice on the current issue?	Given the background in Section 3, the opinion of the ESAC should provide expert advice to ECVAM on three studies that ECVAM conducted in view of assessing whether the three CTA protocols (SHE 6.7; SHE 7.0 and BALB/c) have been sufficiently standardised to be transferable to other laboratories and reproducible between different laboratories and may therefore be fit for future use. In providing this advice, ESAC is requested to take account of the existing information (in particular the OECD DRP) and address also the suitability of the three CTA assays/protocols in question to be used as a basis for the development of OECD test guidelines as foreseen by the OECD in the context of the OECD DRP which led to the present study.

4.2 QUESTION(S) TO BE ADDRESSED

Questions What are the questions and issues that should be addressed in view of achieving the objective of the advice?	The ESAC is requested to address the following three questions: 1) to review whether the study of the three CTAs was conducted appropriately in view of the stated purpose, i.e. of assessing whether the CTA protocols are sufficiently standardised to be transferable and reproducible. In particular the following issues should be addressed: 1. Clarity of the definition of the study objective. 2. Appropriateness of the study design (e.g. chemical selection, number of chemicals used, number of laboratories, acceptance criteria). 3. Appropriateness of the study execution (e.g. were there pre-defined acceptance criteria, were these respected? How were exceptions / deviations handled, e.g. retesting?). 4. Appropriateness of the statistical analysis as used in the protocols and for analysing reproducibility. 2) to assess whether the conclusions as presented in the Study Reports by the Validation Management Team are justified by the information generated during the study and whether they are plausible with respect to existing information and current views (e.g. literature), in particular the OECD DRP on CTAs. In particular the following issues should be addressed: a) Provide a qualitative discussion of the study results/deliverables achieved within the limits of this prevalidation study: • Clarity and completeness of the standardised protocol. • Within laboratory reproducibility • Transferability (critical issues and how they were handled)

	• Between laboratory reproducibility b) Provide a clear presentation of the conclusions presented in the study reports c) Evaluate to which extent the conclusions are justified by the study results alone d) Discuss the plausibility of the conclusion in the light of the study results AND existing historical information as available to the EWG (in particular the OECD DRP which led to this study). 3) to express its opinion with regard to the question whether the CTA protocols standardised and evaluated during the study could indeed be recommended to serve as a basis for an OECD test guideline on in vitro carcinogenicity testing. In particular the following issues should be addressed: a) Similarity of the standardised protocols with respect to the historical protocols (provide to the extent possible a direct comparison and discuss the relative importance of any difference identified). b) Other critical issues and gap analysis (what further work may be useful/required). Please provide a rationale for your proposed position.

4.3 TIMELINES

Timelines concerning this request When does ECVAM require the advice?	Timeline	Indication
	Finalised ESAC Opinion required by:	The ESAC opinion should be available latest during the second half of February 2011 (e.g. 20.1.2011). An attempt will be made to finalise the opinion by written procedure. [N.B. Following ESAC33 (12 October 2010) the entry has been revised (original text below in grey/strikethrough): ~~before February 2011, if possible (i.e. before the OECD WNT meeting)~~]
	Request to be presented to ESAC by written procedure (e.g. due to urgency) prior to the next ESAC	NO
	Request to be presented to ESAC at ESAC plenary meeting	October 2010

5. ECVAM PROPOSALS ON HOW TO ADDRESS THE REQUEST WITHIN ESAC

5.1 ECVAM PROPOSAL REGARDING REQUEST-RELATED STRUCTURES REQUIRED

Specific structures required within ESAC to address the request Does the advice require an ESAC working group, an ESAC rapporteur etc.?	Structure(s) required	Required according to ECVAM?
	S1 ESAC Rapporteur	NO
	S2 ESAC Working Group	YES
	S3 Invited Experts	NO
	Ad S3: If yes – list names and affiliations of suggested experts to be invited and specify whether these are member of the EEP	NO
	If other than above (S1-S3):	NO

5.2 DELIVERABLES AS PROPOSED BY ECVAM

Deliverables What deliverables (other than the ESAC opinion) are required for addressing the request?	Title of deliverable other than ESAC opinion	Required?
	D1 ESAC Rapporteur Report and draft opinion	NO
	D2 ESAC Peer Review Report and draft opinion	YES (ECVAM proposal)
	If other than above (D1-D2):	NO

6. LIST OF DOCUMENTS TO BE MADE AVAILABLE TO THE ESAC

Count	Description of document	Available (YES/NO)	File name
1	VMT-study report SHE pH 6.7	YES	1)ER2010-02_SHE6.7.pdf
2	VMT-study report SHE pH 7.0	YES	2)ER2010-02_SHE7.0.pdf
3	VMT-study report Balb/c 3T3	YES	3)ER2010-02_Balb.pdf
4	CTA SHE pH 6.7 photo catalogue	YES	4)ER2010-02_SHE6.7-photo.pdf
5	CTA SHE pH 7.0 photo catalogue	YES	5)ER2010-02_SHE7.0-photo.pdf
6	CTA Balb/c 3T3 photo catalogue	YES	6)ER2010-02_Balb-photo.pdf
7	OECD. Detailed Review Paper on Cell Transformation Assays for Detection of Chemical Carcinogens, OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No. 31 (2007).	YES	7)ER2010-02_OECD-DRPonCTAs.pdf
8	Combes R., Balls M., Curren R., Fischbach M., Fusenig N., Kirkland D., Lasne A., Landolph J., LeBoeuf R., Marquardt H., McCormick J., Mueller L., Rivedal E., Sabbioni E., Tanaka N., Vasseur P. and Yamasaki H. Cell transformation assay as predictors of human carcinogenicity. Alter. Lab. Anim., 27 (1999) 745-67.	YES	8)ER2010-02_ECVAM-WS-Report-on-CTAs.pdf
9	OECD TG 451 on rodent long term carcinogenicity testing	YES	9)ER2010-02_OECD-TG-451.pdf
10	ECVAM Prevalidation Task Force Report 1: The role of prevalidation in the development, validation and acceptance of alternative methods. ATLA 23, 211-217 (1995)	YES	10)ER2010-02_ECVAM-prevalidation.pdf
11	OECD Series on Testing and Assessment Number 3: Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD, Paris, 2005.	YES	11)ER2010-02_OECD-GuidanceDocument.pdf
12	Hartung T., Bremer S., Casati S., Coecke S., Corvi R., Fortaner S., Gribaldo L., Halder M., Hoffmann S., Roi A.J., Prieto P., Sabbioni E., Scott L., Worth A. and Zuang V. A modular approach to the ECVAM principles on test validity. Alter. Lab. Anim., 32 (2004) 467-72.	YES	12)ER2010-02_ECVAM-modular-approach.pdf

7. TERMS OF REFERENCE OF THE ESAC WORKING GROUP

7.1 ESTABLISHMENT OF THE ESAC WORKING GROUP

During its 33rd meeting on 12 October 2010 the ESAC plenary unanimously decided to establish an ESAC Working Group charged with the detailed scientific review of a study on three Cell Transformation (CTA) protocols.

7.2 TITLE OF THE ESAC WORKING GROUP

Full title:
"ESAC Working Group on the scientific review of 3 Cell Transformation Assay (CTA) prevalidation studies (SHE 6.7, SHE 7.0, BALB)".

Abbreviated title:
"ESAC Working Group CTA"

7.3 MANDATE OF THE ESAC WG

The EWG is requested to conduct a scientific review of the ECVAM study concerning three protocols of the Cell Transformation Assay (CTA). The review needs to address the questions put forward to ESAC by ECVAM.

The review should focus on the appropriateness of design and conduct of the study in view of the study objective and should provide an appraisal to which extent the conclusions of the Validation Management Team (VMT) are substantiated by the information generated during the study and how the information generated relates to the scientific background available.

7.4 DELIVERABLE OF THE ESAC WG

The ESAC WG is requested to deliver to the chair of the ESAC and the ESAC Secretariat a detailed ESAC Working Group Report outlining its analyses and conclusions. A reporting template has been appended (Appendix 1) intended to facilitate the drafting of the report.

The conclusions drawn in the report should be based preferably on consensus. If no consensus can be achieved, the report should clearly outline the differences in the appraisals and provide appropriate scientific justifications.

7.5 PROPOSED TIMELINES OF THE ESAC WG

The Secretariat has proposed timelines which should be agreed upon during the first Teleconference (Item 1 in the table):

Item	Proposed date/time	Action	Deliverable
1	Teleconference 5 November 2010, 14:00 CET	Kick-off teleconference to • discuss the mandate, deliverables, timelines, study background • agree on timelines and meeting dates/times (see item2) • distribute (if appropriate) work and agree on further communication (e.g. TCs if required)	• Agreed timelines • Agreed work plan and distribution
2	First ESAC WG meeting in Ispra • Option 1 (preferred): 30.11. – 2.12.2010 (3 days) • Option 2: 6.12 - 7.12.2010 (2 days) • Option 3 (least preferred): 14.12. – 16.12.2010 (3 days)	• Discussions of the relevant material and preliminary analysis and possible conclusion. • Identification of unresolved issues and disagreements • Identification of process to resolve possible disagreements • Further work distribution and communication means (e.g. TCs) • Beginning of drafting process of report	Possibly preliminary versions of • ESAC WG Report
3	Teleconference 10. January 2011, 14:00 CET	Refresher teleconference (if required) to revisit the status of the work, plan what remains to be done before the second meeting.	
4	Second (last) ESAC WG meeting in Ispra • Option 1: 12.1. – 14.1.2011 (3 days) • Option 2: • 19.1. – 21.1.2011 (3 days)	Finalisation of ESAC WG Report	Final versions of • ESAC WG Report
5	Tuesday 25.1.2011	Handover of report to ESAC chair and Secretariat	Final edited versions (ready for distribution to ESAC): • ESAC WG Report

7.6 QUESTIONS WHICH SHOULD BE ADDRESSED BY THE ESAC WG

The ESAC WG is requested to address the three questions posed to the ESAC which have been broken down further in more specific questions by the ESAC chair, the chair of the ESAC WG and the Secretariat (see section 4.2).

When preparing the final ESAC WG report to address these questions, the ESAC WG is requested to use a pre-defined reporting template. This template (see appendix 1) follows ECVAM's modular approach and addresses to which extent the standard information requirements have been addressed by the study. The template allows moreover for addressing the issues specific studies outlined in section 4.2. The Secretariat will provide guidance if necessary.

APPENDIX 1 REPORTING TEMPLATE FOR THE ESAC WG REPORT

The following suggested template follows the ECVAM modular approach and allows at the same time for the description of the analysis and conclusions concerning more specific questions. The template can be used for various types of validation studies (e.g. prospective full studies, retrospective studies, performance-based studies and prevalidation studies). Depending on the study type and the objective of the study, not all sections may be applicable. However, for reasons of consistency and to clearly identify which information requirements have not been sufficiently addressed by a specific study, this template is uniformly used for the evaluation of validation studies.

Text in red is explanatory, not intended to be part of the title.
One section is clearly not applicable to the present CTA study (identified).

1. Data collection
    - 1.1 Information / data sources used (e.g. reference data)
    - 1.2 Search strategy
    - 1.3 Selection criteria applied to the available information

2. Study objective and design
    - 2.1 Clarity of the definition of the study objective
    - 2.2 Analysis of the scientific rationale provided
    - 2.3 Analysis of the regulatory rationale provided
    - 2.4 Appropriateness of the study design
      (selection of test items, number of test items, number of laboratories, retesting in case of unqualified tests etc.)
    - 2.5 Appropriateness of the statistical evaluation
      (independence of statisticians, statistical method)

3. Test definition (Module 1)
    - 3.1 Quality and completeness of the overall test definition
      (test system, protocol, test acceptance criteria etc.)
    - 3.2 Quality of the background provided concerning the purpose of the test method
    - 3.3 Quality of the documentation and completeness of (a) standardised protocols (SOPs) and (b) prediction models

4. Data quality
    - 4.1 Overall quality of the evaluated data
    - 4.2 Sufficiency of the evaluated data in view of the study objective
    - 4.3 Quality of the reference data for evaluating reliability and relevance[4]

5. Test materials
    - 5.1 Sufficiency of the number of evaluated test items in view of the study objective
    - 5.2 Representativeness of the test items with respect to the applicability domain

[4] OECD guidance document Nr. 34 on validation defines relevance as follows: "Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It sis the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of accuracy (concordance) of a test method."

6. Within-laboratory reproducibility (Module 2)

6.1 Assessment of repeatability and reproducibility in the same laboratory

6.2 Conclusion on within-laboratory reproducibility as assessed by the study

7. Transferability (Module 3)

7.1 Quality of design and analysis of the transfer phase

7.2 Conclusion on transferability to a second laboratory as assessed by the study

In particular: where critical issues that may impact on transferability identified or addressed?

8. Between-laboratory reproducibility (Module 4)

8.1 Assessment of reproducibility in different laboratories

8.2 Conclusion on reproducibility as assessed by the study

9. Predictive capacity (Module 5) N.B. Predictive capacity was outside the scope of the study

9.1 Adequacy of the assessment of the predictive capacity in view of the purpose

9.2 Overall relevance (biological relevance and accuracy) of the test method in view of the purpose

10. Applicability domain (Module 6) N.B. Since this study is not a full validation study, the assessment of the applicability domain is rather limited

10.1 Appropriateness of study design to conclude on applicability domain, limitations and exclusions

10.2 Quality of the description of applicability domain, limitations, exclusions

11. Performance standards (Module 7) N.B. Not applicable to the current study.

11.1 Adequacy of the proposed Essential Test Method Components

11.2 Adequacy of the Reference Chemicals

11.3. Adequacy of the defined Accuracy Values

12. Readiness for standardised use

12.1 Assessment of the readiness for regulatory purposes

12.2. Assessment of the readiness for other uses (in house screening etc.)

12.3 Critical aspects impacting on standardised use

12.4 Gap analysis

Identify, if appropriate, gaps in the study design and/or execution that impact on the stated study objective or the conclusions drawn.

13. Other considerations

Please address any other consideration you might have in relation to the proposed approach under this section.

14. Conclusions and recommendation

14.1 Summary of the study results and conclusions

14.2 Extent to which conclusions are justified by the study results alone

14.3 Extent to which conclusions are plausible in the context of existing information

14.4 Recommendations

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

EURL ECVAM RECOMMENDATION

of 14th March 2012

on

three Cell Transformation Assays

using
Syrian Hamster Embryo Cells (SHE)
and the BALB/c 3T3 Mouse Fibroblast Cell Line

for

In Vitro Carcinogenicity Testing[1]

[1] The ESAC Opinion [2011.01] was adopted on 18 February 2011. The draft EURL ECVAM Recommendation was published as a *Call for Comments* on the IHCP Internet Webpage on 7 December 2011 with a deadline for comments of 31 December 2011.

TABLE OF CONTENTS

EURL ECVAM RECOMMENDATION

of 14.03.12 on

three Cell Transformation Assays (CTA)
using
Syrian Hamster Embryo Cells (SHE)
and the BALB/c 3T3 mouse fibroblast cell line

Executive Summary

The three CTA methods aim to predict the carcinogenic potential of chemicals.

EURL ECVAM fully endorses the ESAC Opinion (dated 18.2.2011) on the performance of these 3 cell transformation assays. In addition, EURL ECVAM recommends that:

- A draft OECD Test Guideline for the SHE CTA should be developed and, considering the limited differences between the protocols for the SHE CTA at pH 6.7 and 7.0, both CTAs should be incorporated into a single Test Guideline;

- Any use of the three CTA protocols should include the requirement of appropriate training by applying the 3 photo catalogues to ensure that scoring is as consistent as possible;

- The BALB/c 3T3 CTA protocol should be further used to expand on the reproducibility of the assay and confirm the suitability of the new statistical approach and data interpretation procedure applied. The BALB CTA should from a 3R perspective be considered more appropriate than the SHE CTA since it uses a cell line.

- The performance characteristics of the SHE CTA methods should be further evaluated by analysing existing data, complemented where necessary by generation of new data to conclude on the regulatory usability;

- The use of CTAs has the potential of partial replacement or reduction when used in a weight of evidence approach for hazard identification and risk assessment.

BACKGROUND TO THIS EURL ECVAM RECOMMENDATION

1. Introduction

1) The in vitro CTAs that have been in use for about 40 years model key stages of in vivo carcinogenicity. In 1998, EURL ECVAM held a workshop with the aim to "seek consensus on ways of increasing the use of mammalian CTAs, for fundamental and applied studies in carcinogenesis, and for the regulatory testing of carcinogens" (Combes et al. 1999). The workshop concluded that the CTA tests were promising but required further development, standardisation and verification prior to being proposed for regulatory use. In parallel, the OECD had in 1997 initiated work on a Detailed Review Paper (DRP) on CTAs for the detection of chemical carcinogens. The DRP was finalised in 2007 (OECD, 2007) concluding that the performance of the Syrian Hamster Embryo (SHE) and the BALB/c 3T3 CTAs were sufficiently adequate for being proposed to be developed into OECD Test Guidelines. A major criticism by regulators was the lacking standardization of protocols and objective criteria for the scoring. On the background of these two activities, and following the recommendations of an expert group which met at EURL ECVAM in 2004, EURL ECVAM coordinated a study aiming at the standardisation and subsequent evaluation of three CTA protocols in terms of transferability and reproducibility. The study was intended to complement the retrospective evaluation of the OECD DRP in view of providing prospective data on reliability but addressing predictive capacity only to a limited extent since an extensive body of existing evidence on the predictivity of CTAs was available and summarised in the OECD DRP.

2) The three CTA protocol variants were:

   (a) The SHE CTA performed at pH 6.7 (SHE pH 6.7 CTA);

   (b) the SHE CTA performed at pH 7.0 (SHE pH 7.0 CTA); and,

   (c) the BALB/c 3T3 CTA.

3) After completion of the study and finalisation of the study reports in summer 2010 (EC-ECVAM, 2010a, b, c), EURL ECVAM requested the EURL ECVAM Scientific Advisory Committee (ESAC) at its meeting on 12 October 2010, to provide scientific advice on this study in the form of an ESAC opinion. ESAC established a Working Group (WG), chaired by two of its members and including internationally renowned experts on CTA, and provided the WG with the mandate to review in detail the results compiled in the three reports (EC-ECVAM, 2010d). Both the ESAC WG report (EC-ECVAM, 2011a) and the ESAC opinion (Annex) were adopted by ESAC on 18 February 2011, and made available to the OECD in time for its annual WNT meeting in mid-April 2011.

4) Based on the above mentioned documents (i.e. ESAC opinion, study reports and the ESAC WG report) and other relevant documents, mainly the OECD DRP (OECD, 2007), EURL ECVAM developed the present recommendation. The aim of any EURL ECVAM recommendation is to provide EURL ECVAM's views on the validity of the test method(s) in question in addition to advice on possible regulatory applicability, limitations and proper scientific use of the test methods, and to suggest possible follow-up activities.

2. Test method definition

Basis of the test method

5) For both SHE CTAs (at pH 6.7 and pH 7.0), the test system is based on SHE cells derived from mid-gestation embryos of Syrian golden hamsters. For the BALB/c 3T3 CTA, the test system is

based on an established mouse fibroblast cell line. The three variants of the CTA assay require visual scoring of specific parameters relating to the phenotype and growth pattern of cells. These effects can be visually detected and scored under the microscope. The parameters measured in the SHE and in the BALB/c 3T3 CTAs are the number of transformed colonies and the number of foci formed, respectively.

Biological and mechanistic relevance of the test method

6) In vitro CTAs have been shown to closely model some stages of the multistage process of in vivo carcinogenesis. The phenomenon of morphological transformation is characterised by changes in the behaviour and growth of cultured cells allowing for progression to the next stage in the transformation process, from a normal cell to a fully malignant cell. A minimum of four phenotypic stages appears to be involved in cell transformation (LeBoeuf et al., 1999), which includes:

   (a) a block in cellular differentiation visualised as morphological transformation in the SHE CTA;

   (b) the acquisition of immortality expressed by unlimited lifespan, an aneuploid karyotype and genetic instability;

   (c) the acquisition of tumourigenicity closely associated with the in vitro phenotypes of focus formation, anchorage-independent growth in semi solid agar and autocrine factor production; and,

   (d) full malignancy, when cells are injected into a suitable host.

7) Of particular interest is the fact that the CTA has the potential to detect both genotoxic and non-genotoxic carcinogens. The use of two-stage protocols can allow for the distinction between tumour initiators and tumour promoters (OECD, 2007).

3. Overall performance of the CTAs

Level of standardisation of the test method

8) As a result of the study, well-described standardized protocols are available for the SHE and the BALB/c 3T3 CTAs. Both of the SHE protocols appear to be transferable (at least to experienced laboratories) and reproducible between laboratories following the ESAC opinion (Annex). For the BALB/c 3T3 CTA, this protocol is due to the introduction of a new and rather specialised statistical method and the data interpretation procedure may require further attention, better definition and refinement of the acceptance and assessment criteria (see ESAC opinion). Moreover, for all three variants of the CTA assay, detailed recommendations have been made by the Validation Management Team (VMT) and the ESAC WG to support further standardisation of the protocols (see also the study reports and the ESAC WG report)(ESAC, 2011) and EURL ECVAM supports these recommendations. Importantly, during the EURL ECVAM study, photo catalogues for each variant of the assay were produced demonstrating typical effects in the three individual CTAs, and these catalogues are expected to support consistent scoring of transformed colonies and foci during training and use of the assay. The recommended protocols and photo catalogues are being published in a Special Issue of Mutation Research on Cell Transformation (Corvi and Vanparys, 2012).

9) EURL ECVAM concludes that the current protocols are sufficiently standardized to be recommended for routine use of the CTAs and INVITTOX protocols to which the photo-catalogues will be attached will be prepared.

Reproducibility

10) A formal evaluation of reproducibility on the basis of standardized protocols has been conducted in the context of the EURL ECVAM study. The ESAC agreed with the VMT that, the two SHE protocols yielded results that were concordant between laboratories and hence reproducible for the substances tested. Moreover, the reproducibility observed in the study is considered plausible with respect to existing data on the assays. This is based upon a consideration of the extensive body of existing data produced with protocols which the ESAC WG analysis indicated are appreciably similar and based upon the apparent robustness of the SHE assays (e.g. as reviewed in the OECD DRP). These conclusions are substantiated by the body of knowledge related to these assays. In particular by (i), the reproducibility evaluations of similar protocols as reported in the literature (Isfort et al., 1996) and, (ii) the overall evaluation of the data contained in the OECD DRP (OECD, 2007).

11) In contrast, evidence from the study supporting reproducibility of the results between laboratories for the BALB/c 3T3 protocol was considered insufficient, as suggested by the need to refine assay assessment criteria and to repeat some experiments to obtain concordant results across laboratories. Also the body of evidence in terms of available data is considerably smaller. It is recommended that the refined BALB/c 3T3 protocol is used in the future to confirm the reproducibility of the assay.

12) According to the ESAC, for the three CTAs evaluated, the confidence in the within-laboratory reproducibility was not sufficiently established because of the use of a single compound which was tested coded and non-coded and was further used as the positive control (Annex). Comparing data produced in different labs with the standardized protocols will help to clarify this issue, in particular if standardized data repository formats are applied. The EURL ECVAM study showed that with adequate training and by using photo catalogues of typical transformation images, a sufficient degree of consistency is reached.

13) EURL ECVAM concludes that data should be collected from the CTAs to enable further verification of their performance. EURL ECVAM will therefore attach standardized data reporting templates to the INVITTOX protocols that are in preparation.

Transferability

14) In general, the proposed test method can be performed in a laboratory that is experienced in routine cell culture techniques (Annex). Considering the nature of the readout (visual scoring), correct scoring of transformed colonies or foci is critical. In the case of the SHE protocols the successful transfer was further supported by the good between-laboratory reproducibility achieved by laboratories that had some experience with CTAs, but not necessarily with the protocol variant considered.

15) EURL ECVAM concludes that transferability should not be a problem for laboratories with sufficient expertise in cell culture. However, the standardized protocols should be strictly followed and sufficient training, in particular for correct scoring is essential. The photo catalogues should be used during training and subsequent routine use.

4. Suggested regulatory use of the CTA test methods

Present and past use

16) As part of its safety assessment process, submitters have in the past provided to the US-FDA (Food and Drug Administration) results from SHE CTA testing as part of the data submission

package. Such results were considered by FDA as supplemental information in its overall product evaluation (Jacobson-Kram and Jacobs, 2005). However, regulatory agencies in general have been reluctant to unconditionally adopt such assays in their routine safety testing schemes, especially as a full replacement for in vivo carcinogenicity testing (OECD, 2009), due, for the most part, to the lack of formal validation data of such assays. Furthermore, one of the main concerns has been the lack of objective criteria to identify and score transformed colonies and foci which could affect the reliability of the test. However, the CTAs are currently being used by academia, the chemical, agro-chemical, cosmetic, pharmaceutical and tobacco industries, and CRO's. Some current uses of the CTAs include:

(a) to provide useful ancillary information when the biological significance of the bioassay result is uncertain (e.g. in pharmaceutical industry);

(b) to clarify in vitro genotoxic positive results by weight of evidence (e.g. in chemical and cosmetic industries);

(c) to screen for non-genotoxic carcinogens (e.g. in agro-chemical industry);

(d) to demonstrate differences and similarities across a chemical class (e.g. in chemical companies within REACH);

(e) to screen for efficacy of chemopreventive agents (in pharmaceutical industry);

(f) to investigate tumor promotion activity (e.g. in agro-chemical and chemical industries); and

(g) for mechanistic studies of carcinogenicity (e.g. in academia and industry).

17) The CTAs are also used to evaluate certain classes of chemicals that have a low predictive capacity in the traditional in vitro genotoxicity tests (e.g. in chemical and cosmetic industries), like the use of the SHE pH 6.7 CTA for testing aromatic amines.

Possible regulatory use

18) Due to the complexity of the events leading to the final adverse effect and based on current opinion, no single in vitro method can provide sufficient information for an unequivocal assessment of the carcinogenicity potential of a substance to satisfy regulatory requirements fully. The CTAs may however provide useful information about possible genotoxic and non-genotoxic carcinogenicity potential for use in conjunction with other data to generate supporting information for hazard identification and risk assessment. The assay may thus be used for these purposes in the context of a weight of evidence approach. Depending on the regulatory context and the extent of other information available from non-testing and testing approaches, it is conceivable that information on the transforming potential of chemicals generated with the CTA may be sufficient for decision-making and may thus in specific cases allow waiving the use of the rodent bioassay. In other cases, the CTA may provide testing data that still require confirmatory testing.

19) The possible use of the SHE and BALB/c 3T3 CTAs for regulatory purposes is mentioned in various recent testing strategies including the FDA guidance for integration of genetic toxicology study results for pharmaceuticals (FDA, 2006) and the guidance on information requirements and chemical safety assessment for REACH (ECHA, 2008). Further possible uses of the SHE CTAs are mentioned in the Scientific Committee on Consumer Products (SCCP)'s notes of guidance for

testing For Testing of Cosmetic Ingredients (SCCP 2010), and in the guidance for testing cosmetics (Pfuhler et al., 2010).

20) However, for use of other classes of chemicals than those so far evaluated, it is recommended to verify that the assay is suitable for that specific application (e.g. testing some reference chemicals of interest). Data from the OECD DRP (OECD, 2007) refer to pure chemicals and show that CTAs can be applied to organic and inorganic chemicals and that they can be used to identify genotoxic and non-genotoxic rodent carcinogens. It is plausible that CTAs can be applied to nanoparticles (Ponti et al., 2009). While there is no sufficient evidence on the performance of the assays using mixtures and formulations, there are no scientific reasons to exclude a priori that the CTAs are not suitable assays to test chemical mixtures and formulations (Breheny et al., 2005).

21) EURL ECVAM concludes that the CTAs have a promising potential for regulatory use, as the study has addressed two major regulatory concerns, i.e. scoring and protocol standardization. However, before concrete regulatory use beyond weight of evidence approach can be recommended, quality controlled historical data shall be used together with new data generated with the standardised protocols for further establishing the capacity of the CTAs to predict the outcome of the rodent bioassay.

Impact on the three Rs

22) The use of the CTAs has the potential to lead to partial replacement and reduction of animal tests (mainly life-time cancer bioassays, OECD 2009) in the regulatory and non-regulatory context. In research, CTAs are, and can be used for research targeting the biological mechanism underlying carcinogenicity. In the regulatory and risk assessment context, the results produced by the CTAs, when considered in conjunction with other available data, may allow concluding on the absence or presence of a carcinogenic chemical hazard. Therefore, high quality CTA information may allow waiving the need to conduct the cancer bioassay. For the SHE CTAs, the use of primary cells from Syrian hamster embryos using pregnant female hamsters may be considered sensitive and appropriate methods of humane killing need to be applied (as outlined in the protocols).

23) EURL ECVAM concludes that the CTAs have a potentially significant 3R impact, as partial replacement to the rodent bioassay, however, it should be noted that from a 3R perspective the BALB CTA is considerably more appropriate since it uses a cell line and not primary embryonic hamster cells, as is the case for the SHE CTA.

5. Limitations

24) There are no known apparent limitations related to specific classes of chemicals that can be tested with the CTAs (OECD, 2007). Implementation and routine use of the CTAs can be limited by the following factors:

Applicability domain:

- The CTA works for pure chemicals falling into the range of the chemicals used in the EURL ECVAM study and included in the OECD DRP. Outside of this range it is recommended to run some suitable reference chemicals to ensure that the CTA can be used.

- There are no available data for the use of mixtures but it is plausible that the CTAs may be applicable for this use as well. In any case, it is essential to ensure that the test chemical reaches the cells.

Practical aspects:

- Cost: the CTA is a rather costly in vitro test (12-35 k€ per substance), but cheap in comparison with the rodent bioassay (1-1.5 M € per substance).

- Throughput: The CTA requires 2-7 weeks per substance, i.e. it has a low throughput. However, this has to be compared with the 3 years a rodent bioassay requires.

- Complexity: the CTA requires high skills with regard to handling of numerous cell plates simultaneously for a relatively long time period, and in particular scoring. Training and the use of the photo catalogues are essential for overcoming these potential limitations.

- X-ray: for the SHE CTAs, the need for X-ray exposed feeder cells to support the growth of target cells requires the access of an irradiation facility.

- The statistical method proposed for the BALB/c 3T3 CTA is not widely used and requires a certain level of expertise and appropriate IT tools.

6. Follow-up activities recommended by EURL ECVAM

- Prepare and publish INVITTOX protocols with the photo catalogues and the data repository template attached, EURL ECVAM has already embarked on this.

- Collection of high quality historical data for retrospective validation, whenever possible supported by new high-quality data generated with the standardised protocols, suggested to be followed-up by the OECD CTA Expert Group.

- Preparations of a combined draft OECD Test Guideline for the two SHE protocols.

- Research and development should be promoted for: (i) human cell based CTAs; (ii), elucidating tumour promoting mechanisms as manifest in CTAs; and, (iii), increasing throughput and reliability of CTAs, e.g. by automation of the visual scoring, if possible together with cost-reduction.

References

Breheny D, Zhang H, Massey ED (2005), Application of a two-stage Syrian hamster embryo cell transformation assay to cigarette smoke particulate matter. Mutat. Res., 572: 45-57.

Combes R, Balls M, Curren R, Fischbach M, Fusenig N, Kirkland D, Lasne A, Landolph J, LeBoeuf R, Marquardt H, McCormick J, Mueller L, Rivedal E, Sabbioni E, Tanaka N, Vasseur P and Yamasaki H (1999), Cell transformation assay as predictors of human carcinogenicity. Alter. Lab. Anim., 27: 745-67.

Corvi, R. and P. Vanparys, (2012) Special Issue on Cell Transformation assays (Eds. Corvi, R and Vanpyris, P.). Mutagen Research: Genetic Toxicology and Environmental Mutagenicity, In Press.

EC-ECVAM (2010a), Report of the Validation Management Team on the ECVAM prevalidation study concerning the SHE pH 6.7 Cell Transformation Assay.

EC-ECVAM (2010b), Report of the Validation Management Team on the ECVAM prevalidation study concerning the SHE pH 7.0 Cell Transformation Assay.

EC-ECVAM (2010c), Report of the Validation Management Team on the ECVAM prevalidation study concerning the BALB/c 3T3 Cell Transformation Assay.

EC-ECVAM (2010d), ECVAM request for ESAC advice on an ECVAM-coordinated prevalidation study concerning the protocols of three Cell Transformation Assays (CTAs) for in vitro carcinogenicity testing. [ESAC request 2010-02]

EC-ECVAM (2011a), ESAC Working Group peer review consensus report on an ECVAM-coordinated prevalidation study concerning the protocols of three Cell Transformation Assays (CTAs) for in vitro carcinogenicity testing. [1102-18 ESAC WG CTA]

ECHA (2008), REACH Guidance on information requirements and chemical safety assessment, Chapter R.7a: Endpoint specific guidance. See chapter R7.7.1 Mutagenicity testing. pp. 374-402. Available at: http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r7a_en.pdf?vers=20_08_08

FDA (2006) Guidance for industry and review staff. Recommended approaches to integration of genetic toxicology study results, http://www.fda.gov/ohrms/dockets/98fr/04d-0493-gdl0002.pdf

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.

Jacobson-Kram D, Jacobs A (2005), Use of genotoxicity data to support clinical trials or positive genetox findings on a candidate pharmaceutical or impurity…now what? Int. J. Toxicol., 24: 129-34.

LeBoeuf RA, Kerckaert KA, Aardema MJ, Isfort RJ (1999), Use of Syrian hamster embryo and BALB/c 3T3 cell transformation for assessing the carcinogenic potential of chemicals. IARC Sci Publ., **146:** 409-25.

OECD (2007), Detailed review on cell transformation assays for detection of chemical carcinogens, OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No. 31. Available at: [www.oecd.org/env/testguidelines]

OECD (2009), Carcinogenicity Studies. Test Guideline No. 451, OECD Guidelines for the Testing of Chemicals, OECD, Paris. Available at: [www.oecd.org/env/testguidelines]

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, Schellauf 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:** 315-24.

Ponti J, Sabbioni E, Munaro B, Broggi F, Marmorato P, Franchini F, Colognato R, Rossi F (2009), Genotoxicity and morphological transformation induced by cobalt nanoparticles and cobalt chloride: an in vitro study in Balb/3T3 mouse fibroblasts. Mutagenesis, 24**:** 439-45.

SCCS European Commission. Health and Consumer Protection, Directorate-General for Health and Consumers. Scientific Committee on Consumer Safety (2010) "The SCCS's Notes of Guidance For Testing of Cosmetic Ingredients and their Safety Evaluation. 7th Revision. Adopted by the SCCS during the 9th plenary meeting of 14 December 2010. SCCS/1416/11. http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_s_004.pdf

ANNEX

ESAC OPINION

based on the ESAC Peer Review
of a EURL ECVAM-coordinated validation study

of

three Cell Transformation Assay (CTA) protocols
for in vitro carcinogenicity testing

ESAC Opinion Nr.	2011-01
Relevant ESAC request Nr.	2010-02
Date of opinion	18.2.2011

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The potential carcinogenicity of chemicals, pharmaceuticals and food additives is a toxicity effect of great concern. To date, the standard approach to assess carcinogenicity for regulatory purposes is the two year bioassay in rodents (OECD TG 451). Several in vitro alternative methods have been developed. While in vitro genotoxicity tests address only induction of genetic damage as a mechanisms leading to carcinogenicity, in vitro Cell Transformation Assays (CTAs) have been shown to recapitulate stages of in vivo carcinogenesis. Exposure of cultured cells to carcinogenic substances in the CTA can lead to cell transformation involving changes in cell behaviour/phenotype (e.g. proliferation control, altered cell morphology, changed colony growth patterns, anchorage independent growth). Transformed cells can lead to tumour formation in vivo when injected in a suitable host, underlining the biological relevance of the CTAs for carcinogenicity testing.

Continuing previous evaluations of the potential utility of the CTAs for standardised applications including regulatory testing (Combes et al., 1997), EURL ECVAM conducted, from 2005 to 2010, a prevalidation study on three protocol variants of the CTA. Two of the three protocols were based on cells from Syrian Hamster Embryos (the 'SHE' variant of the CTA), the "SHE pH 6.7" and "SHE pH 7.0" assays. One protocol was based on the BALB/c 3T3 cell line, the "BALB/c assay". The study addressed the three aspects of prevalidation (EC- ECVAM, 1995): protocol refinement, transfer and preliminary assessment of, within the limits of a small scale study, protocol reproducibility within and between laboratories.

Following a request from EURL ECVAM to ESAC in October 2010 (EC-ECVAM 2010d; c.f. Annex 2) for scientific advice on this study, the ESAC set up a Working Group (ESAC WG) charged with the detailed scientific peer review of this prevalidation study.

After careful peer review of the study reports (EC-ECVAM 2010 a-c) and considering the detailed ESAC WG peer review consensus report (EC-ECVAM 2011), the ESAC concludes, in agreement with the Validation Management Team (VMT), that the reliability of the BALB/c protocol was not adequately addressed in the present study.

In contrast, in case of the SHE pH 6.7 and SHE pH7.0, the study data indicate that sufficiently standardised protocols have been produced which appear transferable. While the data relating to the assessment of within-laboratory reproducibility were considered insufficient, the data indicated acceptable between-laboratory reproducibility for the compounds tested and when considering the type of study (i.e. prevalidation).

In view of the possible standardised use of the SHE protocols including for regulatory purposes, the development of a common SHE protocol is recommended describing both pH variants (i.e. pH6.7 and pH7.0). In a next step, test performance needs to be characterised on the basis of a larger set of chemicals covering a broad range of chemical classes and mechanisms of action. This should in particular include the evaluation of more data for non-carcinogens.

However, when planning future activities, the extent to which existing testing information on such chemicals could be used to describe SHE protocol performance for a specific purpose should be considered carefully. Such information may either be published or reside with relevant stakeholders. In the opinion of the ESAC it is conceivable and plausible, considering the extensive body of information available, that historical SHE testing data could be used to arrive at a robust characterisation of the SHE test method performance to support possible standardised use, including for regulatory purposes.

This view is based, firstly, on the apparent robustness of the SHE assays as demonstrated by an analysis of published data in the OECD DRP: the predictions compiled in this report were obtained using non-standardised protocols and showed nevertheless a high degree of concordance. Secondly, an analysis carried out by the ESAC WG (cf. Annex 2 of ESAC WG Report) indicated appreciable

similarity of the historical SHE protocols and the standardised protocols in this study, supporting the possible integration of prospective with existing testing data.

Finally, the ESAC recommends, that future activities towards the possible use of these assays should start with the definition of the intended purpose which is expected to facilitate a detailed and targeted characterisation of test performance (e.g. predictive capacity, limitations) on the basis of new or existing information.

1. Mandate of the ESAC

On its meeting on 12 October 2010, the ESAC was requested by EURL ECVAM (see Annex 2) to conduct a scientific review of an EURL ECVAM-coordinated prevalidation study on three protocols of the Cell Transformation Assay (CTA) for carcinogenicity testing ("CTA prevalidation study"). Two of the three protocols were based on cells from Syrian Hamster Embryos (the 'SHE' variant of the CTA). The two protocols differed mainly with respect to the pH of the medium in which the cells are kept (either pH 6.7 or pH 7.0) and the protocols are hereunder referred to as "SHE pH 6.7" and "SHE pH 7.0" assays. One protocol was based on the BALB/c 3T3 cell line and is referred to hereunder as "BALB/c assay". The study addressed the three aspects of prevalidation (ECVAM 1995): protocol refinement, transfer and preliminary assessment, on the basis of a small scale study, of protocol reproducibility within and between laboratories.

1.1 General objective of the advice to be given by ESAC

Given the background made available in Section 3 of the associated request of EURL ECVAM to ESAC (EC-ECVAM 2010d; Annex 2) and all documentation made available to the ESAC (EC-ECVAM 2010d; Annex 2), the opinion of the ESAC should provide expert advice to EURL ECVAM on a prevalidation study that EURL ECVAM conducted in view of assessing whether three protocols of the Cell Transformation Assay (the variants were the SHE pH6.7, the SHE pH7.0 and BALB/c protocols) have been sufficiently standardised to be transferable to other laboratories and reproducible between different laboratories and may therefore be fit for future use.

In providing this advice, ESAC is requested to take account of the existing information, in particular the OECD DRP (OECD, 2007) and address also the suitability of the three CTA assays/protocols in question to be used as a basis for the development of OECD test guidelines as foreseen by the OECD in the context of the OECD DRP which led to the present study.

1.2 Questions that should be addressed by the ESAC and its Working Group

The specific questions related to this mandate are listed in section 4.2 of Annex 2 (EC-ECVAM 2010d).

1.3 Background to the ESAC Mandate

1.3.1 Background to the study

It is important to note that this EURL ECVAM study, performed from 2005 to 2010, was planned on the background of past EURL ECVAM activities towards the possible use of CTAs for carcinogenicity testing (Combes et al., 1997) but also on the background of the, at the time ongoing, OECD project towards the drafting of a Detailed Review Paper (DRP) on historical CTA data (the "OECD DRP") which took place from 1997 to 2007.

Both projects therefore temporally overlapped to some extent and information on progress in both projects was mutually taken into consideration: while the EURL ECVAM studies relied to a great

extent on reference data compiled in the OECD DRP, the recommendations made in the OECD DRP took already into account possible results of the prevalidation study conducted by EURL ECVAM at that time and aiming at the refinement of selected CTA protocols.

According to the recommendations of the OECD DRP (published in 2007 while the EURL ECVAM study was still ongoing), the present studies (if successful) should contribute to decisions regarding the incorporation of the CTA assays into an OECD test guideline / guidelines. However, the specific purpose of the tests within the framework of an OECD test guideline was not defined in the OECD DRP.

Since the OECD DRP was to provide detailed information on the predictive capacity of the CTAS, the EURL ECVAM study was planned as a complement to this DRP. Consequently, the study focused on the development and evaluation (i.e. transfer/reproducibility) of standardised and well documented protocols that could serve as a possible basis for an OECD test guideline. In contrast, predictive capacity, which typically is addressed to a preliminary extent during prevalidation studies, was not defined as a study objective and there is hence no description of the assays' accuracy (sensitivity/specificity) in the study reports.

However, since test items with reference data of the rodent bioassay (and in part of IARC) were used to assess reproducibility (see figure 1 of ESAC WG report), the predictive capacity of the standardised protocols could be calculated on the basis of the test items assessed (see Annex 1 of the ESAC WG report).

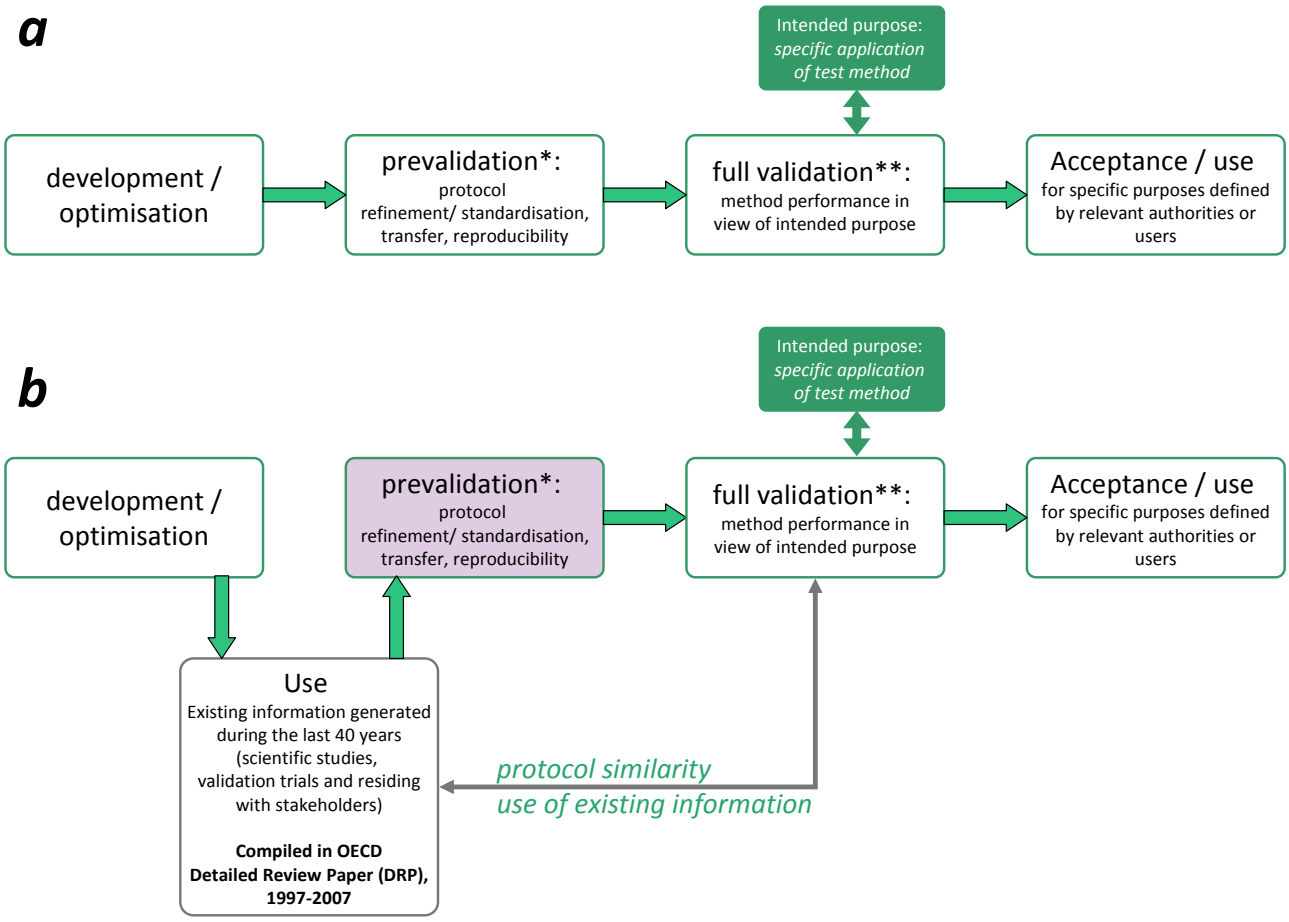
1.3.2 Specifics of this study

From the background to the study outlined in 1.3.1, it is obvious that the current prevalidation study represents a specific case (see figure 1).

Typically, a prevalidation study follows the development / optimisation phase of a new test method (figure 1a). Prevalidation studies are designed to further refine and standardise the test method's protocol (Standard Operating Procedure, SOP) and transfer the SOP from a lead laboratory to other laboratories for an assessment of between-laboratory reproducibility on the basis of a limited set of test items. This is done in order to determine whether the new method and in particular its associated protocol (SOP) is mature and robust enough to merit progressing into a costly and time-consuming full validation study which aims at a performance characterisation of the assay in view of a specific purpose.

In contrast, the present prevalidation study was planned on the background of an extensive body of data, compiled in the OECD DRP (figure 1b). The aim of the prevalidation study was, as for all prevalidation studies, the standardisation of protocols, their transfer to other laboratories and subsequent assessment of protocol reproducibility. However, in view of full test performance characterisation, the existing information and the extent to which it can be used, is an important issue and has been touched on in the recommendations (section 3.15) of this document. An important aspect in this context is the similarity of historical protocols versus protocols as generated during this study. This issue has been addressed, in a preliminary manner, by this review and is presented in more detail in Annex 2 of the ESAC WG report.

Figure 1: Validation flow. (a) Typical flow from development over prevalidation, validation to acceptance and use. **(b)** The situation for the CTAs. The assays have been in use for considerable time and a large body of data has been produced on the basis of the CTAs. The extent to which existing information can be used for validation and performance characterisation needs to be carefully considered.



*) Upon completion, information requirements (modular information) only partly fulfilled.
**) Upon completion, all information requirements are satisfactorily fulfilled in view of purpose

2. Summary of the ESAC Opinion

Taking into account (a) the detailed review of the ESAC WG including the WG's analysis concerning the similarity of existing protocols and those generated during the prevalidation study (EC-ECVAM 2011), (b) the information made available to ESAC by EURL ECVAM including the Validation Study Reports (EC-ECVAM 2010 a-c), (c) the EURL ECVAM request for ESAC advice outlining the ESAC's mandate (EC-ECVAM 2010d; c.f. see Annex 2) the ESAC has the following opinion:

(1) The reliability of the BALB/c protocol was not adequately addressed in the present study. Major concerns are, inter alia, that repeat testing was executed without blinding and also that the 'assay assessment criteria' (allowing translating the measurements into predictions on the transforming potency of substances) require further refinement as already suggested by the VMT coordinating the study. The ESAC appreciates why further refinements were made to the BALB/c protocol as a consequence of testing during the study, but believes that the final protocol having undergone these modifications should be tested in future trials.

(2) Despite shortcomings in study design and execution (see point 6), the study data indicate acceptable reliability for both the SHE pH 6.7 and SHE pH 7.0 protocols for the compounds tested and when considering the type of study (i.e. prevalidation). Reproducibility was assessed by analysing

the concordance of predictions made in the participating laboratories and when comparing the predictions to in vivo carcinogenicity reference data from respected sources reviewed in the OECD Detailed Review Paper (OECD, 2007) including from IARC (IARC 2009), the National Toxicology Program (NTP database available online) and Gold and Zeiger (Gold & Zeiger 1997). Moreover, successful transfer can be concluded from the data on between-laboratory reproducibility.

(3) Although the number of test items used and consequently the chemical domain occupied is rather limited due to a shortcoming of study design (see paragraph 6) and the nature of the study (prevalidation), the ESAC is nevertheless of the opinion that it is plausible that this level of reproducibility would extend to other chemical domains as well. This notion is supported by (a) the apparent similarity of the historical versus the current protocols (c.f. analysis of protocol similarity performed by the ESAC WG, EC-ECVAM 2011) as well as (b) the apparent robustness of the CTA assays in general (see point 6 for more details). Therefore, the SHE assay protocols as standardised during this study are at least sufficiently reproducible, for those chemicals tested, to be considered for use in a regulatory setting.

(4) Despite the small number of items tested (see paragraph 6), these nevertheless reflect a certain range of the possible combinations concerning carcinogenicity and genotoxicity profiles[2]. Briefly, in case of the SHE assays, 4/6 substances tested are carcinogens (benzo(a)pyrene, 2,4-diaminotoluene, o-toluidine, 3-Methylcholantrene), while 2/6 of the substances are non-carcinogens (anthracene and phthalic anhydride). 2/4 of the carcinogenic substances are clearly genotoxic in vivo and in vitro assays (benzo(a)pyrene and 2,4-diaminotoluene), while for one the overall evidence suggests that it is a genotoxic carcinogen despite some inconclusive in vivo genotoxicity data (3-methylcholanthrene)[3]. The remaining substance (o-toluidine HCL) has equivocal data from in vivo and in vitro genotoxicity tests and could be regarded as a non-genotoxic carcinogen.

(5) While reproducibility was promising for the SHE assays, note should be taken of the fact that robust conclusions on reproducibility cannot be drawn from this prevalidation study alone. The dataset generated during this study is too small to allow sufficient characterisation of key items of test method performance. These include: reproducibility on the basis of a larger and different set of chemicals (including weakly transforming agents), predictive capacity, applicability and possible limitations.

(6) The ESAC noted that there were some shortcomings with regard to study design and execution. Although these did not critically influence the outcome of the study, they nevertheless may constitute deviations from what may be seen as good practice in validation according to EURL ECVAM's approach and International guidance (OECD 2005). These issues include the planning and execution of the selection of reference data (see 3.4), the low number of test items (n=6), even when considering that this is a prevalidation study, the within-laboratory phase (see 3.6), the transferability phase (see 3.7) and the lack of clear provisions for retesting (e.g. number of admissible retests in case of data not fulfilling test acceptance criteria).

(7) In view of possible recommendations concerning future activities towards application of these assays for standardised testing purposes, the ESAC has the following opinion: the next step following on from this prevalidation exercise normally would be a full prospective validation study to characterise the performance of the test methods in view of standardised use including possible regulatory use. Such a study would comprise a set of test substances covering a wide range of chemical classes / possible mechanisms of action (e.g. genotoxic/non-genotoxic) and which is large enough for a statistical evaluation of predictions into two (dichotomous) classes: transforming or

[2] A description of these mechanism-related issues relating specifically to chemical selection can be found in Vinken et al (2008).

[3] Evidence for the genotoxicity of 3-Methylcholantrene is provided for instance in: Moorthy et al., 2007; Xu et al, 2005; Rihn et al., 2000; Moorthy et al., 1993; Bryla & Wyand, 1992; for bibliographic references see section 15.

non-transforming agents. In line with the OECD DRP it is noted by the ESAC that also pharmaceuticals should be included in future analyses of test performance. The ESAC recommends extending this to food additives (e.g. flavours, fragrants and other food supplements).

(8) However, when planning future activities, it should be carefully considered to which extent existing information can be used. In the opinion of the ESAC it is conceivable and plausible that historical SHE testing data could be used to arrive at a robust characterisation of the SHE test method performance to support possible regulatory use. This view is based, firstly, on the apparent robustness of the SHE assays as demonstrated by an analysis of published data in the OECD DRP: the predictions compiled in this report were obtained using non-standardised protocols and showed nevertheless a high degree of concordance. Secondly, an analysis carried out by the ESAC WG (EC-ECVAM 2011: Annex 2) indicated appreciable similarity of the historical SHE protocols and the standardised protocols in this study supporting the possible integration of prospective with existing test data. About 500 coded and un-coded compounds have up to now been tested using the SHE assay by many laboratories. Careful use and reanalysis of these historical data, which include validation studies, may be able to supplement or substitute for a new full prospective validation study of the SHE assay. The ESAC WG draws attention to two publications that highlight the good predictive capacity of the SHE assays (Isfort et al, 1996; Mauthe et al, 2001).

(9) The ESAC is of the opinion, that any further activities towards assay performance characterisation should pay attention to the chemical selection and, if using existing information, to an appropriate description of the toxicity potency of substances, their physicochemical properties, chemical class and mechanism of action in order to define applicability and, in particular, possible limitations of the assays. The ESAC has two concerns with respect to the information compiled in the OECD DRP (OECD, 2007): (a) the lack of explicitness regarding the completeness of the existing data presented and whether data selection criteria based on study quality were defined and had been applied and (b) the lack of a description of the transforming potency of the chemicals analysed. Reproducibility may have been overestimated if most data are based on transforming / non-transforming agents that have generally shown unequivocal results in the past (i.e. no record of discordant results between laboratories). Thus, future activities using existing information should address these issues before new prospective studies are planned and/or before drawing conclusions on test performance on the basis of existing data. Finally, the ESAC recommends, that future activities towards the possible use of these assays should start with the definition of the intended purpose which is expected to facilitate a detailed and targeted characterisation of test performance (e.g. predictive capacity, limitations) on the basis of new or existing information.

3. Detailed opinion of the ESAC

The following paragraphs follow largely the same structure as used in the ESAC WG report.

3.1 Data collection

3.1.1 Reference Data

All reference data used in the prevalidation study are derived from the Detailed Review Paper on Cell Transformation Assays for Detection of Chemical Carcinogens published by the OECD in 2007 (hereunder abbreviated as "OECD DRP").

The current prevalidation studies make use of data from six of the chemicals reported in the OECD DRP as reference data to assess reproducibility as measured through the **concordance of predictions within and between laboratories** and in reference to in vivo carcinogenicity classifications as

published in the OECD DRP, which in turn – in case of these chemicals – refers to classifications by the International Agency for Research on Cancer (IARC), Gold and Zeiger (1997) and the U.S. National Toxicology Program (NTP) database.

Briefly, the predictions generated by the Cell Transformation Assay (CTA) protocols in this study allow classification of the test chemical either as a "transforming agent" or as a "non-transforming agent" (for SHE, based on calculation of "Morphological Transformation Frequency", MTF; for BALB/c 3T3, based on measurement of the number of type III foci). Predictions (transforming / non-transforming agent) obtained in different laboratories were assessed for consistency (concordance) between the laboratories and were compared with in vivo carcinogenicity data as reported in the OECD DRP as the "reference standard".

Moreover, relevant EURL ECVAM workshop reports and recent research papers related to (a) the scoring of observed effects, (b) the mechanistic understanding and (c) between-laboratory reproducibility are discussed in the prevalidation study reports (sections 1.4 to 1.7 pp.10-13 in all three reports), but have not been used as reference data.

For additional details, see section 1.1 of ESAC WG report

3.1.2 Search strategy to retrieve reference data associated with the test items

The ESAC working group (ESAC WG) noted that there was apparently no detailed search strategy established for identifying suitable reference data. However, taking into account that this was a small scale study which did not attempt to define the predictive capacity or the applicability domain of the three CTAs studied, but focused on protocol refinement and reproducibility, this fact was not considered relevant in this context.

For additional details, see section 1.2 of ESAC WG report

3.1.3 Selection criteria for reference data

The ESAC WG noted that there was apparently no detailed set of selection criteria established to reject/accept retrieved data. The OECD DRP was taken as reliable source although it is not clearly described in the OECD DRP how the quality of the data had been controlled.

For additional details, see section 1.3 of ESAC WG report

3.2 Study objective

3.2.1 Clarity of the study objective

The objective of the studies was considered clear and comprehensive: standardisation of CTA protocols and subsequent assessment of these protocols for reproducibility and transferability.

For additional details, see section 2.1 of ESAC WG report

3.2.2 Intended scientific rationale

The intended scientific rationale was explained as far as our current understanding of the cellular mechanisms involved in carcinogenesis (primarily in rodent cells) allows. The reported prevalidation study does not contribute to this scientific understanding, but builds upon evidence (provided primarily by the OECD DRP) that genotoxic as well as non-genotoxic carcinogens induce cell transformation in SHE and BALB/c 3T3 cells while non-carcinogenic substances do not.

For additional details, see section 2.2 of ESAC WG report.

3.2.3 Regulatory rationale

The regulatory rationale remains somewhat open although it is acknowledged by the ESAC WG that even screening data and supportive data within a Weight of Evidence framework can be used for regulatory purposes and may thus constitute a "regulatory rationale". However, recommendations for a more precise definition of the regulatory usability of these tests should have been made in the reports since such use is mentioned as one of the motives for the study (see also Section 15. Recommendations). For additional details, see section 2.3 of ESAC WG report

3.2.4 Appropriateness of study design

Overall, the study design was considered appropriate for assessing the reproducibility and transferability of the standardised protocols, despite shortcomings relating to the design/planning of (1) the test item selection (even when considering that this is a prevalidation study, there are concerns regarding the number representativeness of test chemicals with regard to chemical class and mechanism of action), (2) the within-laboratory variability phase and (3) the transferability phase.

For additional details, see section 2.4 of ESAC WG report

3.2.5 Appropriateness of statistical evaluation

The statistical evaluation of the test data generated during the study appears appropriate. However, the methods of statistical analysis used in the test method procedures (SOP) and the assay assessment criteria need a critical revision.

For additional details, see section 2.5 of ESAC WG report

3.3 Test definition

Overall the tests were adequately defined considering the objective of this study. The overall purpose of the study (development of OECD guidelines) was clear, but the specific purpose of the tests was neither defined by OECD nor by the VMT. Validation is the assessment of the satisfactory performance of a system designed for a specific purpose. Considering this, the absence of a clearer definition of the purpose or possible use of the tests may have influenced the study design, e.g. with respect to the test chemical selection. It is therefore recommended that the intended purpose is sufficiently considered when planning a validation study including a prevalidation exercise. The SOPs were found acceptable provided some minor revisions, including the ones recommended by the VMT.

For additional details, see section 3. of ESAC WG report

3.4 Data quality

3.4.1 Quality of the evaluated data

In general, the data quality was good. Acceptance criteria are broad enough to anticipate different outcomes of the assays when applied properly. Discrepancies were explained.

3.4.2 Sufficiency of the evaluated data in view of the study objective

The data generated and evaluated did not allow for either a proper assessment of within-laboratory reproducibility or the success of the test transfer within the context of a dedicated study phase (transferability) for all three assays.

In contrast, the data produced for assessing between-laboratory reproducibility were considered sufficient for the SHE assays and may moreover be used to infer the success of transfer. The lack of appropriate testing for both modules - within-laboratory reproducibility and transferability - is, however, not considered compliant with standard practice in validation.

3.4.3 Quality of the reference data

The quality of the reference data was assumed sufficient for assessing reproducibility, being based on the OECD DRP and well-regarded sources (i.e. IARC, Gold & Zeiger and NTP database). However, it was noted that there were apparently no provision for assessing the quality of data reported in the OECD DRP and consequently in the present study.

For additional details on data quality, see section 4. of ESAC WG report

3.5 Test items

3.5.1 Sufficiency of the number of evaluated test items in view of the study objective

The number of chemicals (n=6) is judged to be sufficient, with respect to statistical requirements, to assess reproducibility (the main study objective).

However, although it is acknowledged that this is not a full validation study, the number of substances tested is low. More specifically, the number appears low to adequately cover, also for the purposes of a prevalidation study, the range of possible types of chemicals in view of the most prominent underlying mechanism of action (i.e. genotoxic / non-genotoxic) for an endpoint as complex as carcinogenicity.

Thus, the reproducibility assessment is restricted in this case to substances that belong to the same chemical classes (i.e. organic substances; inorganic compounds have not been tested) and which have the same mechanism of action as the ones tested in the prevalidation study. Considering the more advanced SHE assays, 3/6 of the substances were clear genotoxic carcinogens, only 1/6 of the substances was a possible non–genotoxic carcinogen (more details in 3.5.2).

For additional details, see section 5.1 of ESAC WG report

3.5.2 Representativeness of the test items with respect to the applicability domain

It is not the objective of this study to assess the limitations of the tests. However, based on the ESAC WG's analysis (cf. section 12.1; Annex 2) showing the apparent similarity of the historical protocols compared with the protocols from this prevalidation study, it is conceivable that historical data as reported in the OECD DRP could help describe the applicability domain of the test methods in the future.

When considering the test items of this study, it appears that mainly clear positives (transforming agents) and clear negatives (non-transforming agents) have been tested, but no "equivocal" substances (known to be able to lead to discordant results within and between laboratories) which may have challenged the reproducibility of the protocols more adequately.

For the two SHE protocols, a few materials, such as reserpine, cinnamyl anthranilate, or ethylene thiourea, which have given discordant results in previous interlaboratory studies (Tu et al. 1986; Jones et al. 1988) might have helped better define the transferability and reproducibility of these newly standardised protocols.

When considering the complexity of the endpoint (i.e. regarding decisions on "genotoxic/non-genotoxic" and "carcinogenic/non-carcinogenic") as well as considering the small number of test items (n=6), the test items covered a range of the possible combinations of (non)genotoxic and

(non)carcinogenic (cf. ESAC WG report section 5.2, Box 1 'categorisation tree' and Box 2 reproducing table 35, p68 in the SHE pH6.7 assay).

Briefly, in case of the SHE assays, 4/6 substances tested are carcinogens. These are benzo(a)pyrene, 2,4-diaminotoluene, o-toluidine and 3-methylcholanthrene. 2/6 are non-carcinogens when considering reference data from the rodent bioassay (anthracene and phthalic anhydride). 1 of these non-carcinogens (anthracene) is currently not classifiable according to IARC.

Furthermore, 2/4 carcinogenic substances studied are clearly genotoxic in vivo and in vitro assays (benzo(a)pyrene and 2,4-diaminotoluene), while for one the overall evidence suggests that it is a genotoxic carcinogen despite some inconclusive in vivo genotoxicity data (3-methylcholanthrene). The remaining substance (o-toluidine HCL) has equivocal data from in vivo and in vitro genotoxicity tests and could be regarded as a non-genotoxic substance.

As the concept of non-genotoxic carcinogens has only been accepted rather recently and many substances found to be carcinogens have been tested repeatedly for genotoxicity, it is possible that such substances with equivocal genotoxicity data could be regarded as non-genotoxic carcinogens (e.g. o-toluidine).

For additional details, see section 5.2 of ESAC WG report

3.6 Within laboratory reproducibility

The ESAC WG feels that within-laboratory reproducibility was not clearly established due to inadequate study design: only one chemical was tested. Moreover, this substance was the positive control (benzo(a)pyrene for the SHE CTAs, 3-methylcholanthrene for the BALB/c 3T3 CTA). These may, due to their strong transforming potency, lead to an overestimation of reproducibility. While, within laboratory reproducibility for this single substance was, not surprisingly, high in the SHE and BALB/c assays and also between-laboratory reproducibility was good, one chemical only (in addition the PC) cannot be regarded as a sufficient dataset to conclude on this module in compliance with good validation practice.

For additional details, see section 6. of ESAC WG report

3.7 Transfer phase / Transferability

The transfer phase was adequately described and appropriately executed so allowing proper test method conduct in the other laboratories for the subsequent analysis of between laboratory reproducibility. However, how the success of the transfer was assessed and what criteria were used to judge the transfer successful was not clearly described. Moreover, ease of transferability was not assessed through the testing of test items. While this is not a prerequisite for prevalidation studies, the current study nevertheless did not fully address one of its objectives (i.e. assessment of the transferability module).

The success of the transfer programme was not demonstrated in separate experiments. All participating laboratories had some experience with CTAs. Thus, the ease of transferability to a laboratory without any CTA experience was not demonstrated. However, this is in any case not a formal requirement for a prevalidation study (OECD guidance document Nr. 34), and in this case it is noted that successful transfer may be inferred from the good between-laboratory reproducibility.

For additional details, see section 7. of ESAC WG report

3.8 Between-laboratory reproducibility

The final outcome following the implementation of the assessment criteria was considered reproducible for the SHE assays. For the BALB/c 3T3 assay, refinement of the assessment criteria is required.

Between-laboratory reproducibility was assessed through analysis of the concordance of predictions for the six test substances obtained by the involved laboratories. The predictions concerned the classification of test substances as potential transforming agents / non-transforming agents in the CTA assays. The CTA predictions were compared with the reference data associated with the test chemicals. These data are in vivo carcinogenicity predictions taken from the OECD DRP report which, for the test chemicals, are based on IARC classifications, the Gold & Zeiger and the NTP databases.

Based on the data generated and reported, the ESAC WG agrees with the VMT that the two SHE protocols yield results which are concordant between laboratories and hence reproducible for the substances tested.

In contrast, evidence supporting reproducibility of the results between laboratories for the BALB/c 3T3 protocol was considered insufficient, as suggested by the need to refine assay assessment criteria and to repeat some experiments to obtain concordant results across the laboratories.

For additional details, see section 8. of ESAC WG report

3.9 Predictive capacity

Although predictive capacity was outside the scope of the study objective, it is noteworthy that the predictions made by the SHE assays for the six chemicals were in most cases correct (6/6 corrections were correct in the SHE pH7.0, while 5/6 were correct in the SHE pH6.7). While the chemicals selected may have a bias towards reproducible results (clear negatives and strong positives), the results are nevertheless reassuring and add to the database of CTA testing data. However, based on the ESAC WG's analysis (cf. section 12.1; Annex 2) showing apparent similarity of the historical protocols with the protocols from the prevalidation study, it is conceivable that historical data as reported in the OECD DRP could help describe the predictive capacity of the SHE test methods in the future.

3.10 Applicability and limitations of the test methods

Since this study is not a full validation study, the assessment of the applicability domain is rather limited. However, based on the ESAC WG's preliminary analysis (cf. Annex 2 of ESAC WG report) showing apparent similarity of the historical protocols with the protocols from the prevalidation study, it is conceivable that historical data as reported in the OECD DRP could help describe the applicability and limitations of the SHE test methods in the future.

3.11 Performance Standards

Not applicable to this study.

3.12 Readiness for standardised use

3.12.1 Readiness for regulatory use

The data generated during this prevalidation study, when viewed on their own, are insufficient to draw conclusions on readiness for regulatory use of the SHE assay protocols, although these showed acceptable reproducibility.

However, an analysis by the ESAC WG identified considerable similarity in the historical SHE protocols and the standardised protocols in this prevalidation study. This supports the view that a substantial amount of the existing testing information from the SHE assays could be used for future considerations on their performance (e.g. predictive capacity, applicability / limitations) required to define their regulatory utility.

This view is further supported by the apparent robustness of the SHE assays as demonstrated in the OECD DRP: the data show considerable concordance with regard to the predictions made even though these older protocols may have differed to some extent and clearly no standardised test procedures had been used. These predictions are moreover relevant when compared with in vivo carcinogenicity data derived from respected sources (e.g. IARC, NTP database).

The ESAC WG, therefore, believes that future activities aimed at more precise definition of test method performance of the SHE assay and possible regulatory utility of the associated SHE protocols can be based on both, prospective testing but also on the analysis of existing historical information (e.g. a meta-analysis using defined search and data selection criteria based on study quality).

The ESAC WG notes, based upon current opinion, that no single method can provide sufficient information for an unequivocal assessment of the carcinogenicity potential of a substance to satisfy regulatory requirements fully. The SHE assays may provide information about possible genotoxic and non-genotoxic carcinogens for use in conjunction with other data (e.g. in the context of a "weight-of-evidence" approach). Some recommendations on possible approaches towards the expansion of the performance characterisation of these methods are made in section 3.15, notwithstanding the fact that the specific regulatory use needs to be defined by the relevant authorities for the purpose in mind.

The study results show that, in contrast to the SHE data, the BALB/c 3T3 protocol still requires optimisation (concerning for example the assessment criteria for the assay) and is at present neither ready to enter full validation nor consideration for regulatory use based on existing information.

For additional details, see section 12.1 of ESAC WG report

3.12.2. Assessment of the readiness for other uses

The ESAC considers the CTAs useful for testing compounds belonging to the same class of chemicals as those used in the reported prevalidation studies (screening purposes) and to generate supporting information for hazard identification and risk assessment (weight of evidence). Moreover, the CTAs will continue to be useful also for mechanistic studies of the transformation process.

For additional details, see section 12.2 of ESAC WG report

3.12.3 Critical aspects impacting on standardised use

The performance characteristics of the SHE methods need to be carefully analysed through prospective testing and/or analysis of existing information (protocol similarity supports the use of historical data) before the SHE protocols can be used in standardised applications (regulatory or non-regulatory). This analysis should include a careful examination of the chemical classes tested. Moreover, some improvement of the SHE protocols such as the development of common protocol for the two pH variants and a better description of some of the protocol steps (cell preparation) should be performed before standardised use is considered.

Concerning the BALB/c 3T3 CTA, further optimisation of the protocol is needed. These modifications, including those suggested by the VMT, should be tested in further trials before standardised use is considered.

For additional details, see section 12.3 of ESAC WG report

3.13 Other considerations

Detailed suggestions on prevalidation conduct and test method SOPs and their use have been made in the ESAC WG report (section 15.).

3.14 Conclusions on the study

As a scientific piece of work the study is impressive and succeeded in generating, in case of the SHE assays, standardised protocols including associated photo catalogues to support consistent scoring. Sufficient between-laboratory reproducibility was demonstrated for these standardised protocols. Moreover, the predictions yielded were in most cases relevant when compared to reference data (rodent bioassay and IARC class, where available).

Despite some shortcomings in study design (mainly with respect to the number of items tested, but also the design of the within-laboratory reproducibility requirement), the study succeeded with respect to its stated goals. The extent to which the various information requirements of this prevalidation study were addressed and fulfilled in view of the objective of the study is summarised in table 1.

For additional details, see section 14 of ESAC WG report

Table 1: Extent to which information requirements were addressed and fulfilled in view of the objective of the prevalidation study

	Protocol standardisation	**Within-laboratory reproducibility**	**Transferability**	**Between-laboratory reproducibility**
SHE pH 6.7	**Achieved**. Single reporting format would have been beneficial. Development of the photo catalogue is considered a major merit of the study.	**Not sufficiently addressed** Only one substance tested. In some cases this was the PC (intrinsic propensity to generate reproducible results). Study design of WLR phase was very variable.	**Successfully transferred to experienced laboratories.** Success of transfer was not tested empirically but can be deduced from information on BLR.	**Satisfactorily demonstrated** Satisfactory for the substances tested*
SHE pH 7.0	**Achieved** Single reporting format would have been beneficial. Development of the photo catalogue is considered a major merit of the study.	**Not sufficiently addressed** Only one substance tested. In some cases this was the PC (intrinsic propensity to generate reproducible results). Study design of WLR phase was very variable.	**Successfully transferred to experienced laboratories.** Success of transfer was not tested empirically but can be deduced from information on BLR.	**Satisfactorily demonstrated** Satisfactory for the substances tested*
BALB/c 3T3	**Not finalised** Assessment criteria were insufficient at outset of study. Further definition suggested by VMT and ESAC WG. These improvements need now to be assessed by testing.	**Not sufficiently addressed** Only one substance tested. In some cases this was the PC (intrinsic propensity to generate reproducible results). Study design of WLR phase was very variable.	**Successfully transferred to experienced laboratories.** Success of transfer was not tested empirically but can be deduced from information on BLR.	**Promising but insufficient** Further refinement of the test method required.

3.15 Recommendations for future work in view of standardised use of the test methods

As briefly outlined in section 1.3, the specific purpose of the CTAs within the framework of, for example, an OECD test guideline was neither defined in the OECD DRP nor by the Validation Management Team (VMT) when planning the current study. This has implications on the future strategy regarding test performance characterisation and possible (regulatory) use of the assays.

Test performance characteristics of a test method (e.g. applicability, limitations, predictive capacity) are to some extent dependent on the specific intended purpose of a test method. Thus, in the absence of a clear intended purpose (e.g. 'use within a test strategy for industrial chemicals to identify positives to waive confirmatory in vivo testing'), the characterisation of test performance remains difficult or on a general level which does not reflect the needs and constraints of possible applications.

Vice versa, the intended regulatory use of test method is easier to define if a precise description of test performance is available. This mutual interdependency of the test method performance characterisation and test method purpose description may, if not addressed in a forward-looking manner, hamper implementation of test methods that appear to be promising for standardised applications including regulatory testing.

In order to avoid such a situation the ESAC recommends that future activities aiming at the potential use of the SHE assays (e.g. OECD test guideline development) should commence with a definition of the intended use of the assays based on the information available (e.g. the current study and the OECD DRP). This will allow a targeted and in-depth description of test performance for the purpose in mind. A strategy towards test method characterisation has been outlined below, using existing information to the extent possible.

3.15.1. Recommendations for the SHE assays

Although the present study succeeded in generating standardised protocols which appear reproducible, the SHE assays are at present not yet ready for regulatory use.

In any case, a revision of the protocols with the aim of incorporating the two SHE cell protocols into one single protocol describing both pH variants (pH6.7 and pH7.0) should be considered. As a minimum, the two protocols should be harmonized as much as possible. Moreover, considering the nature of the readout (visual scoring), it is recommended that the SOPs contain a specific subsection on training and transfer of the assays to naïve laboratories[4]. The definition of proficiency chemicals would support such transfer and help laboratories to assess whether they are capable of conducting the assay.

More importantly, the assays require still a complete description of their performance on the basis of a considerably larger set of chemicals including, if necessary for the envisaged purpose, pharmaceuticals and food additives. Future test substances should include substances that challenge the transferability and reproducibility (i.e. substances with discordant results between laboratories) as well as substances representing a range of possible mechanisms of action. Such performance characterisation should include information on (a) predictive capacity, (b) applicability and, more importantly, limitations of the assays, (c) reproducibility, as well as (d) ease of transferability.

When planning future steps of performance characterisation, the extent to which historical data (including earlier validation studies) can be taken into account, should be carefully considered, as these data could supplement or even substitute for a full new validation study of the SHE assays.

[4] The word 'naive laboratories' in the context of validation refers to laboratories that are inexperienced with regard to the use of a specific test method, i.e. they have not conducted this method or variants of it before.

Moreover, the extent to which prospective testing is required to fully characterise the SHE assays will depend on (a) the intended purpose of the assays including a more precise concept concerning their possible regulatory use and (b) the information that may have – in the meantime – become available in the literature.

The following strategy is recommended in order to gain more robust information towards a complete test performance characterisation of the SHE cell assays especially for regulatory purposes:

STEP 1 – Analysis of existing information:

Any future activity towards the standardised / regulatory use of the SHE method should start with a critical analysis of the considerable body of existing testing information (either published or residing with stakeholders).

It is conceivable that, after analysis of the historical datasets (e.g. with respect to chemical class, mechanism of action, carcinogenic potency) test performance can be satisfactorily described through retrospective validation and meta-analysis of data alone, without further need for prospective testing.

Importantly, such an analysis should also go back to original data and not only rely on processed data such as contained in the OECD DRP. Moreover, an evidence-based approach should be employed using a predefined search strategy for retrieving all relevant information and minimum acceptance criteria for data quality.

Should this analysis show that there are gaps in the existing data sets (e.g. with regard to chemical classes, transforming potency), STEP 2 or STEP 3 should be considered.

STEP 2 – Targeted prospective testing of gap substances:

Should the retrospective evaluation of existing information performed in STEP 1 not suffice for a satisfactory description of test performance in view of the intended purpose, a small and targeted prospective study should be conducted providing information on assay performance for those "gap substances" identified in STEP 1. The testing information generated during STEP 2 may then supplement the existing information compiled in STEP 1.

STEP 3 – Full prospective validation:

Should the information generated during STEP 1 and/or STEP2 not suffice for the intended purpose, a full prospective validation study should be conducted using the SHE protocol(s) produced during this prevalidation study but taking into account the improvements of the SOPs as suggested by the EURL ECVAM/ESAC.

3.15.2. Recommendations for the BALB/c 3T3 assay

The BALB/c 3T3 assay is at present and following this prevalidation study not yet ready for regulatory use requires further optimisation, including refinement of the acceptance and assessment criteria.

However, considering the specificities of the BALB/c 3T3 assay (e.g. use of a well established cell line, no feeder cells needed so no irradiation facility required) compared to the SHE assays, further use of the refined protocol is encouraged to expand the data on assay reproducibility and the appropriateness of the assay assessment criteria (including statistical methodology used) for generating relevant predictions. These steps should precede a more complete test performance characterisation which may follow the same strategy as outlined for the SHE assays (see 3.15.1).

For additional details, e.g. concerning suggestions for improvement of the SOPs of the SHE and BALB/c CTAs, see section 15. of ESAC WG report.

4. References

Bryla P; Weyand EH (1992) Detection of PAH:DNA Adducts from Auto-oxidation Using 32P-Postlabeling. Cancer Letters, Vol. 65, No. 1: 35-41, 25

Combes R, Balls M, Curren R, Fischbach M, Fusenig N, Kirkland D, Lasne A, Landolph J, LeBoeuf R, Marquardt H, McCormick J, Mueller L, Rivedal E, Sabbioni E, Tanaka N, Vasseur P and Yamasaki H (1999) Cell transformation assay as predictors of human carcinogenicity. Alter. Lab. Anim., 27 745-67.

EC-ECVAM (1995) ECVAM Prevalidation Task Force Report 1: The role of prevalidation in the development, validation and acceptance of alternative methods. ATLA 23, 211-217

EC-ECVAM (2010a) Report of the Validation Management Team on the ECVAM prevalidation study concerning the SHE pH6.7 Cell Transformation Assay.

EC-ECVAM (2010b) Report of the Validation Management Team on the ECVAM prevalidation study concerning the SHE pH7.0 CTA.

EC-ECVAM (2010c) Report of the Validation Management Team on the ECVAM prevalidation study concerning the BALB/c 3T3 CTA.

EC-ECVAM (2010d) ECVAM request for ESAC advice on an ECVAM-coordinated prevalidation study concerning the protocols of three Cell Transformation Assays (CTA) for in vitro carcinogenicity testing. [ESAC request 2010-02]

EC-ECVAM (2011) ESAC Working Group Consensus report on an ECVAM-coordinated prevalidation study concerning the protocols of three Cell Transformation Assays (CTA) for in vitro carcinogenicity testing.

Gold LS and Zeiger E, Eds. (1997) Handbook of carcinogenicity and genotoxicity databases. CRC Press, Boca Raton (FL), USA 754 pp.

IARC (2009) Agents reviewed by the IARC Monographs Volumes 1-100A (alphabetical order). http://monographs.iarc.fr/ENG/Classification/Listagentsalphorder.pdf

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.

Jones CA, Huberman E, Callaham MF, Tu A, Hallowell W, Pallota S, Sivak A, Lubet RA, Avery MD and Kouri RE (1988) An interlaboratory evaluation of the Syrian hamster embryo cell transformation assay using eighteen coded chemicals. Toxicol. In Vitro, 2**:** 103-116.

Mauthe RJ, Gibson DT, Bunch RT and Custer L (2001) The Syrian Hamster Embryo (SHE) Cell Transformation Assay: Review of the Methods and Results. Toxicol Pathol. 29 Suppl:138-46.

Moorthy B; Chen S; Li D; Randerath K (1993) 3-Methylcholanthrene-inducible liver cytochrome(s) P450 in female Sprague-Dawley rats: possible link between P450 turnover and formation of DNA adducts and I-compounds. Carcinogenesis. 14(5):879-86.

Moorthy B; Muthiah K; Fazili IS; Kondraganti SR; Wang L; Couroucli XI; Jiang W (2007) 3-Methylcholanthrene elicits DNA adduct formation in the CYP1A1 promoter region and attenuates reporter gene expression in rat H4IIE cells. Biochem Biophys Res Commun., Mar 23; 354(4):1071-7.

NTP, NTP website at http://www.ntp-server.niehs.nih.gov.

OECD (2005) Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD Series on testing and assessment Nr. 34.

OECD (2007) Detailed review paper on cell transformation assays for detection of chemical carcinogens, OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No. 31. http://appli1.oecd.org/olis/2007doc.nsf/linkto/env-jm-mono(2007)18.

Rihn BH; Bottin MC; Coulais C; Rouget R; Monhoven N; Baranowski W; Edorh A; Keith G (2000) Genotoxicity of 3-methylcholanthrene in liver of transgenic big Blue mice. Environ Mol Mutagen.; 36(4):266-73.

Lu LJ (1990) Transplacental DNA damage: effects of chemical types, gestation stage, Ah Locus, and species. J Cancer Res Clin Oncol;116(Suppl Pt 2):1002

Tu A, Hallowell W, Pallotta S, Sivak A, Lubet RA, Curren RD, Avery MD, Jones C, Sedita BA, Huberman E, Tennant RW, Spalding J and Kouri RE (1986) An interlaboratory comparison of transformation in Syrian hamster embryo cells with model and coded chemicals. Environ. Mutagen., 8**:** 77-98.

Vinken M, Doktorova T, Ellinger-Ziegelbauer H, Ahr HJ, Lock E, Carmichael P, Roggen E, van Delft J, Kleinjans J, Castell J, Bort R, Donato T, Ryan M, Corvi R, Keun H, Ebbels T, Athersuch T, Sansone SA, Rocca-Serra P, Stierum R, Jennings P, Pfaller W, Gmuender H, Vanhaecke T and Rogiers V (2008) The carcinoGENOMICS project: critical selection of model compounds for the development of omics-based in vitro carcinogenicity screening assays, Mutat. Res., 659**:** 202-10.

Xu M; Nelson GB; Moore JE; McCoy TP; Dai J; Manderville RA; Ross JA; Miller MS (2005) Induction of Cyp1a1 and Cyp1b1 and formation of DNA adducts in C57BL/6, Balb/c, and F1 mice following in utero exposure to 3-methylcholanthrene. Toxicol Appl Pharmacol. 209(1**):**28-38

Annex 1 of the ESAC Opinion

ESAC and ESAC Working Group charged with the scientific review

EURL ECVAM Scientific Advisory Committee
Dr. Nathalie ALÉPÉE
Dr. David BASKETTER
Dr. Neil CARMICHAEL
Prof. Jacques R. CHRETIEN
Prof. Lucio G. COSTA
Dr. Rodger CURREN
Prof. A. Wallace HAYES
Prof. Coenraad HENDRIKSEN
Dr. Dagmar JÍROVÁ
Prof. Walter PFALLER
Dr. Erwin ROGGEN
Prof. Vera ROGIERS
Dr. Andrea SEILER
Prof. Kristin SCHIRMER
Prof. Ruud A. WOUTERSEN

ESAC Working Group CTA
Dr. Erwin ROGGEN (ESAC member, Chair of the WG)
Dr. Rodger CURREN (ESAC member)
Dr. David LOVELL (Invited expert, EEP)
Dr. Edgar RIVEDAL (Invited expert, EEP)
Dr. Takeki TSUTSUI (Invited expert following an ICATM proposal from JaCVAM; EEP)

ESAC Secretariat
Dr. Claudius GRIESINGER
Dr. Pascal PHRAKONKHAM (specific support)

Annex 2 of the ESAC Opinion

EURL ECVAM request to ESAC for scientific advice concerning the EURL ECVAM-coordinated CTA prevalidation study

ESAC Request 2010-02

EURL ECVAM Scientific Advisory Committee

(ESAC)

EURL ECVAM REQUEST FOR ESAC ADVICE

on an EURL ECVAM-coordinated prevalidation study concerning the protocols of three Cell Transformation Assays (CTA) for carcinogenicity testing

Title page information	
Abbreviated title of ESAC request	ESAC peer review of and ESAC opinion on the EURL ECVAM-led prevalidation study of three cell transformation assays for carcinogenicity testing: 1) SHE pH 6.7 assay 2) SHE pH 7.0 assay 3) Balb/c 3T3 assay.
ESAC REQUEST Nr.	2010-02
Filename	ESAC REQUEST_2010 02_CTA+ESAC-WG-Mandate-approved.doc
Template used for preparing request	EP 2.01
Date of finalising request	2010-09-14
Date of submitting request to ESAC	2010-10-02
Request discussed through	Plenary at ESAC33 at 2010/10/12
Opinion expected at (date)	Through written procedure in January 2011 (before OECD WNT in March)

TABLE OF CONTENTS

1. TYPE OF REQUEST

Request Type		**Identify request ("YES")**
R1 **ESAC Peer Review of a Prevalidation Study or Validation Study**		**YES**
If R1)applies please specify further:		
►Prevalidation Study	**YES** The study is a complement to the recommendations of the OECD Detailed Review Paper on Cell Transformation Assays. The study addressed protocol standardisation, transferability and reproducibility (but not performance) of three protocols of cell transformation assays in view of establishing standardised protocols for future consistent use, e.g. through the development of OECD test guidelines for in vitro carcinogenicity testing.	
►Prospective Validation Study		
►Retrospective Validation Study		
►Validation Study based on Performance Standards		
R2 Scientific Advice on a test method submitted to EURL ECVAM for validation (e.g. the test method's biological relevance etc.)		
R3 Other Scientific Advice (e.g. on test methods, their use; on technical issues such as cell culturing, stem cells etc.)		

2. TITLE OF STUDY OR PROJECT FOR WHICH SCIENTIFIC ADVICE OF THE ESAC IS REQUESTED

Prevalidation of three cell transformation assays for carcinogenicity testing:

1) SHE pH 6.7 assay
2) SHE pH 7.0 assay
3) Balb/c 3T3 assay

3. BRIEF DESCRIPTION OF THE STUDY OR PROJECT

1) Background to carcinogenicity testing and available alternative methods

The potential for a compound to induce carcinogenicity is a crucial consideration when establishing hazard and risk assessment of chemicals and pharmaceuticals in humans. To date, the standard approach to assess carcinogenicity at a regulatory level is the 2-year bioassay in rodents (OECD TG 451; Ref. 1).

Several in vitro alternatives have been developed for predicting carcinogenicity. Of these, the in vitro genotoxicity tests address only one mechanism involved in carcinogenicity, the induction of genetic damage. In contrast, in vitro Cell Transformation Assays (CTAs) have been shown to involve a multistage process that closely models some stages of in vivo carcinogenesis: CTAs can detect phenotypic changes of cultured cells as a result of exposure to test materials (i.e. chemicals, products etc.). These phenotypic/morphological changes are a result of the transformation of cultured cells which involves changes in cell behaviour and proliferation control (e.g. altered cell morphology, changed colony growth patterns and anchorage –independent growth). Moreover, transformed cells can evolve to be tumorigenic when injected in a suitable host. Importantly, CTAs are to date the only optimised tests that allow the detection of both genotoxic and non-genotoxic carcinogens. CTAs have been in use for about 40 years and are currently being performed by academia, the chemical, agro-chemical, cosmetic and pharmaceutical industries. CTAs are conducted in-house as well as at contract research organisations to screen for potential carcinogenicity as well as investigate mechanisms of carcinogenicity. While CTAs are currently not used routinely for regulatory testing, they are frequently used for internal (in-house) safety assessment of chemicals, drugs, etc. and are considered to provide additional useful information to the prevailing tests that are used for assessing carcinogenic potential.

2) The OECD Detailed Review Paper as the basis for this prevalidation study

In order to systematically assess the performance of the CTAs, the Organisation for Economic Co-operation and Development (OECD) finalised in 2007 a "Detailed Review Paper on Cell Transformation Assays For Detection of Chemical Carcinogens" (OECD DRP). The OECD DRP focused on the analysis of the predictive capacity (relevance) of three CTAs and addressed also some elements of reliability: (1) the Syrian hamster embryo (SHE) assay, (2) the BALB/c 3T3 assay and (3) the C3H10T1/2 assay. A substantial body of existing and published data was evaluated (SHE n=264 chemicals; BALB/c 3T3 n=184; C3H10T1/2 n=141). The OECD DRP concluded that the performances of two of the assays, the SHE assay and Balb/c 3T3 assay, were sufficiently adequate and should be developed into formal OECD test guidelines (OECD DRP, Ref. 2). Further, the same OECD DRP recommended that although considerable data on the performance of the assays were available, a formal assessment of the assays, in particular focusing on development of a standardised transferable and reproducible protocol, would be important for preparation of such OECD test guidelines.

3) Study objectives and design

Based on the OECD DRP and several EURL ECVAM expert meetings (Combes et al., 1999, Ref. 3), EURL ECVAM initiated a study on the two CTAs found most relevant by the OECD DRP on the basis of the available information, the SHE and the BALB/c 3T3 assays. The study constitutes a complement to the extensive OECD DRP and its conclusions. In agreement with the conclusions of the OECD DRP, EURL ECVAM focused on the development and evaluation of standardised, well-documented protocols that could serve as a basis for an OECD test guideline. In summary, the study was organised and designed taking into account:

- the **objective of the study** to address protocol standardisation and an assessment of transferability and reproducibility of the standardised CTA protocols but not their predictive capacity (which is addressed by the OECD DRP) and
- the **high costs and considerable time** required to perform the assays as well as the limited funding and resources which could be made available by EURL ECVAM.

The study addressed the three classical aspects of Prevalidation: I) protocol refinement/standardisation; II) protocol transfer and III) protocol performance (ECVAM 1995, Ref.4; OECD guidance document on validation, 2005, Ref. 5). With respect to the modular approach of validation (Hartung et al., 2004, Ref. 6), the study assessed information concerning module 1) test

definition, module 2) within-laboratory reproducibility, module 3) transferability, module 4) between-laboratory reproducibility.

The study addressed three variants of CTA protocols: two SHE protocol variants (cells at pH 6.7 and at pH 7.0, respectively) and the CTA based on the Balb/c 3T3 A31 cell line. Each protocol was assessed using six chemicals. In contrast to the Balb/c 3T3 protocol which required more substantial refinement, both SHE protocols were already available in the literature and results of these have been reported in the OECD DRP. Between-laboratory reproducibility was examined in three laboratories except for the SHE 7.0 protocol, where four laboratories were involved.

4) Results and Conclusions

The Validation Management Team (VMT) concluded that, for the SHE pH 6.7 and the SHE pH 7.0 CTAs, the study had demonstrated that standardised protocols were available which could be the basis for future use. These protocols and the assay system itself have been shown to be transferable between laboratories, and are reproducible within- and between-laboratories. For the Balb/c 3T3 method, an improved protocol has been developed, which allowed obtaining reproducible results. However, further testing of the improved Balb/c 3T3 protocol is recommended (see Validation Study Reports, Ref. 7-9). Moreover, the VMT concluded that the appropriate training and the use of the photo catalogues (see Photo Catalogues, Ref. 10-12) developed during the protocol refinement phase, led to a consistent scoring of transformed colonies and foci.

Overall, these results in combination with the extensive database summarized in the OECD DRP support the utility of in vitro CTAs for the assessment of carcinogenicity potential.

References

1 OECD TG 451 on rodent long term carcinogenicity testing

2 OECD. Detailed Review Paper on Cell Transformation Assays for Detection of Chemical Carcinogens, OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No. 31 (2007).

3 Combes R., Balls M., Curren R., Fischbach M., Fusenig N., Kirkland D., Lasne A., Landolph J., LeBoeuf R., Marquardt H., McCormick J., Mueller L., Rivedal E., Sabbioni E., Tanaka N., Vasseur P. and Yamasaki H. Cell transformation assay as predictors of human carcinogenicity. Alter. Lab. Anim., 27 (1999) 745-67.

4 ECVAM Prevalidation Task Force Report 1: The role of prevalidation in the development, validation and acceptance of alternative methods. ATLA 23, 211-217 (1995)

5 OECD Series on Testing and Assessment Number 3: Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD, Paris, 2005.

6 Hartung T., Bremer S., Casati S., Coecke S., Corvi R., Fortaner S., Gribaldo L., Halder M., Hoffmann S., Roi A.J., Prieto P., Sabbioni E., Scott L., Worth A. and Zuang V. A modular approach to the ECVAM principles on test validity. Alter. Lab. Anim., 32 (2004) 467-72.

7 VMT-study report SHE pH 6.7

8 VMT-study report SHE pH 7.0

9 VMT-study report Balb/c 3T3

10 CTA SHE pH 6.7 photo catalogue

11 CTA SHE pH 7.0 photo catalogue

12	CTA Balb/c 3T3 photo catalogue

4. OBJECTIVES, QUESTIONS, TIMELINES

4.1 OBJECTIVE

Objective Why does EURL ECVAM require advice on the current issue?	Given the background in Section 3, the opinion of the ESAC should provide expert advice to EURL ECVAM on three studies that EURL ECVAM conducted in view of assessing whether the three CTA protocols (SHE 6.7; SHE 7.0 and BALB/c) have been sufficiently standardised to be transferable to other laboratories and reproducible between different laboratories and may therefore be fit for future use. In providing this advice, ESAC is requested to take account of the existing information (in particular the OECD DRP) and address also the suitability of the three CTA assays/protocols in question to be used as a basis for the development of OECD test guidelines as foreseen by the OECD in the context of the OECD DRP which led to the present study.

4.2 QUESTION(S) TO BE ADDRESSED

Questions What are the questions and issues that should be addressed in view of achieving the objective of the advice?	The ESAC is requested to address the following three questions: 1) to review whether the study of the three CTAs was conducted appropriately in view of the stated purpose, i.e. of assessing whether the CTA protocols are sufficiently standardised to be transferable and reproducible. In particular the following issues should be addressed: 1. Clarity of the definition of the study objective. 2. Appropriateness of the study design (e.g. chemical selection, number of chemicals used, number of laboratories, acceptance criteria). 3. Appropriateness of the study execution (e.g. were there pre-defined acceptance criteria, were these respected? How were exceptions / deviations handled, e.g. retesting?). 4. Appropriateness of the statistical analysis as used in the protocols and for analysing reproducibility. 2) to assess whether the conclusions as presented in the Study Reports by the Validation Management Team are justified by the information generated during the study and whether they are plausible with respect to existing information and current views (e.g. literature), in particular the OECD DRP on CTAs. In particular the following issues should be addressed: a) Provide a qualitative discussion of the study results/deliverables achieved within the limits of this prevalidation study: • Clarity and completeness of the standardised protocol. • Within laboratory reproducibility • Transferability (critical issues and how they were

	handled) • Between laboratory reproducibility b) Provide a clear presentation of the conclusions presented in the study reports c) Evaluate to which extent the conclusions are justified by the study results alone d) Discuss the plausibility of the conclusion in the light of the study results AND existing historical information as available to the EWG (in particular the OECD DRP which led to this study). 3) to express its opinion with regard to the question whether the CTA protocols standardised and evaluated during the study could indeed be recommended to serve as a basis for an OECD test guideline on in vitro carcinogenicity testing. In particular the following issues should be addressed: a) Similarity of the standardised protocols with respect to the historical protocols (provide to the extent possible a direct comparison and discuss the relative importance of any difference identified). b) Other critical issues and gap analysis (what further work may be useful/required). Please provide a rationale for your proposed position.

4.3 TIMELINES

Timelines concerning this request When does EURL ECVAM require the advice?	**Timeline**	**Indication**
	Finalised ESAC Opinion required by:	The ESAC opinion should be available latest during the second half of February 2011 (e.g. 20.1.2011). An attempt will be made to finalise the opinion by written procedure.
	Request to be presented to ESAC by written procedure (e.g. due to urgency) prior to the next ESAC	NO
	Request to be presented to ESAC at ESAC plenary meeting	October 2010

5. EURL ECVAM PROPOSALS ON HOW TO ADDRESS THE REQUEST WITHIN ESAC

5.1 EURL ECVAM PROPOSAL REGARDING REQUEST-RELATED STRUCTURES REQUIRED

Specific structures required within ESAC to address the request Does the advice require an ESAC working group, an ESAC rapporteur etc.?	**Structure(s) required**	**Required according to EURL ECVAM?**
	S1 ESAC Rapporteur	NO
	S2 ESAC Working Group	**YES**
	S3 Invited Experts	NO
	Ad S3: If yes – list names and affiliations of suggested experts to be invited and specify whether these are member of the EEP	NO
	If other than above (S1-S3):	NO

5.2 DELIVERABLES AS PROPOSED BY EURL ECVAM

Deliverables What deliverables (other than the ESAC opinion) are required for addressing the request?	**Title of deliverable other than ESAC opinion**	**Required?**
	D1 ESAC Rapporteur Report and draft opinion	NO
	D2 ESAC Peer Review Report and draft opinion	**YES (EURL ECVAM proposal)**
	If other than above (D1-D2):	NO

6. LIST OF DOCUMENTS TO BE MADE AVAILABLE TO THE ESAC

Count	Description of document	Available (YES/NO)	File name
1	VMT-study report SHE pH 6.7	YES	1)ER2010-02_SHE6.7.pdf
2	VMT-study report SHE pH 7.0	YES	2)ER2010-02_SHE7.0.pdf
3	VMT-study report Balb/c 3T3	YES	3)ER2010-02_Balb.pdf

4	CTA SHE pH 6.7 photo catalogue	YES	4)ER2010-02_SHE6.7-photo.pdf
5	CTA SHE pH 7.0 photo catalogue	YES	5)ER2010-02_SHE7.0-photo.pdf
6	CTA Balb/c 3T3 photo catalogue	YES	6)ER2010-02_Balb-photo.pdf
7	OECD. Detailed Review Paper on Cell Transformation Assays for Detection of Chemical Carcinogens, OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No. 31 (2007).	YES	7)ER2010-02_OECD-DRPonCTAs.pdf
8	Combes R., Balls M., Curren R., Fischbach M., Fusenig N., Kirkland D., Lasne A., Landolph J., LeBoeuf R., Marquardt H., McCormick J., Mueller L., Rivedal E., Sabbioni E., Tanaka N., Vasseur P. and Yamasaki H. Cell transformation assay as predictors of human carcinogenicity. Alter. Lab. Anim., 27 (1999) 745-67.	YES	8)ER2010-02_ECVAM-WS-Report-on-CTAs.pdf
9	OECD TG 451 on rodent long term carcinogenicity testing	YES	9)ER2010-02_OECD-TG-451.pdf
10	ECVAM Prevalidation Task Force Report 1: The role of prevalidation in the development, validation and acceptance of alternative methods. ATLA 23, 211-217 (1995)	YES	10)ER2010-02_ECVAM-prevalidation.pdf
11	OECD Series on Testing and Assessment Number 3: Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. OECD, Paris, 2005.	YES	11)ER2010-02_OECD-GuidanceDocument.pdf
12	Hartung T., Bremer S., Casati S., Coecke S., Corvi R., Fortaner S., Gribaldo L., Halder M., Hoffmann S., Roi A.J., Prieto P., Sabbioni E., Scott L., Worth A. and Zuang V. A modular approach to the ECVAM principles on test validity. Alter. Lab. Anim., 32 (2004) 467-72.	YES	12)ER2010-02_ECVAM-modular-approach.pdf

7. TERMS OF REFERENCE OF THE ESAC WORKING GROUP

7.1 ESTABLISHMENT OF THE ESAC WORKING GROUP

During its 33rd meeting on 12 October 2010 the ESAC plenary unanimously decided to establish an ESAC Working Group charged with the detailed scientific review of a study on three Cell Transformation (CTA) protocols.

7.2 TITLE OF THE ESAC WORKING GROUP

Full title:
"ESAC Working Group on the scientific review of 3 Cell Transformation Assay (CTA) prevalidation studies (SHE 6.7, SHE 7.0, BALB)".

Abbreviated title:

"ESAC Working Group CTA"

7.3 MANDATE OF THE ESAC WG

The EWG is requested to conduct a scientific review of the EURL ECVAM study concerning three protocols of the Cell Transformation Assay (CTA). The review needs to address the questions put forward to ESAC by EURL ECVAM.

The review should focus on the appropriateness of design and conduct of the study in view of the study objective and should provide an appraisal to which extent the conclusions of the Validation Management Team (VMT) are substantiated by the information generated during the study and how the information generated relates to the scientific background available.

7.4 DELIVERABLE OF THE ESAC WG

The ESAC WG is requested to deliver to the chair of the ESAC and the ESAC Secretariat a detailed **ESAC Working Group Report** outlining its analyses and conclusions**.** A reporting template has been appended (Appendix 1) intended to facilitate the drafting of the report.

The conclusions drawn in the report should be based preferably on consensus. If no consensus can be achieved, the report should clearly outline the differences in the appraisals and provide appropriate scientific justifications.

7.5 PROPOSED TIMELINES OF THE ESAC WG

The Secretariat has proposed timelines which should be agreed upon during the first Teleconference (Item 1 in the table):

Item	Proposed date/time	Action	Deliverable
1	Teleconference 5 November 2010, 14:00 CET	Kick-off teleconference to • discuss the mandate, deliverables, timelines, study background • agree on timelines and meeting dates/times (see item2) • distribute (if appropriate) work and agree on further communication (e.g. TCs if required)	• Agreed timelines • Agreed work plan and distribution
2	First ESAC WG meeting in Ispra • Option 1 (preferred): 30.11. – 2.12.2010 (3 days) • Option 2: 6.12 - 7.12.2010 (2 days) • Option 3 (least preferred): 14.12. – 16.12.2010 (3 days)	• Discussions of the relevant material and preliminary analysis and possible conclusion. • Identification of unresolved issues and disagreements • Identification of process to resolve possible disagreements • Further work distribution and communication means (e.g. TCs)	Possibly preliminary versions of • ESAC WG Report

		• Beginning of drafting process of report	
3	Teleconference 10. January 2011, 14:00 CET	Refresher teleconference (if required) to revisit the status of the work, plan what remains to be done before the second meeting.	
4	Second (last) ESAC WG meeting in Ispra • Option 1: 12.1. – 14.1.2011 (3 days) • Option 2: • 19.1. – 21.1.2011 (3 days)	Finalisation of ESAC WG Report	Final versions of • ESAC WG Report
5	Tuesday 25.1.2011	Handover of report to ESAC chair and Secretariat	Final edited versions (ready for distribution to ESAC): • ESAC WG Report

7.6 QUESTIONS WHICH SHOULD BE ADDRESSED BY THE ESAC WG

The ESAC WG is requested to address the three questions posed to the ESAC which have been broken down further in more specific questions by the ESAC chair, the chair of the ESAC WG and the Secretariat (see section 4.2).

When preparing the final ESAC WG report to address these questions, the ESAC WG is requested to use a pre-defined reporting template. This template (see appendix 1) follows EURL ECVAM's modular approach and addresses to which extent the standard information requirements have been addressed by the study. The template allows moreover for addressing the issues specific studies outlined in section 4.2. The Secretariat will provide guidance if necessary.

APPENDIX 1

REPORTING TEMPLATE FOR THE ESAC WG REPORT

The following suggested template follows the EURL ECVAM modular approach and allows at the same time for the description of the analysis and conclusions concerning more specific questions. The template can be used for various types of validation studies (e.g. prospective full studies, retrospective studies, performance-based studies and prevalidation studies). Depending on the study type and the objective of the study, not all sections may be applicable. However, for reasons of consistency and to clearly identify which information requirements have not been sufficiently addressed by a specific study, this template is uniformly used for the evaluation of validation studies.

Text in red is explanatory, not intended to be part of the title.
One section is clearly not applicable to the present CTA study (identified).

1. Data collection

1.1 Information / data sources used (e.g. reference data)
1.2 Search strategy
1.3 Selection criteria applied to the available information

2. Study objective and design

2.1 Clarity of the definition of the study objective
2.2 Analysis of the scientific rationale provided
2.3 Analysis of the regulatory rationale provided
2.4 Appropriateness of the study design
(selection of test items, number of test items, number of laboratories, retesting in case of unqualified tests etc.)
2.5 Appropriateness of the statistical evaluation
(independence of statisticians, statistical method)

3. Test definition (Module 1)

3.1 Quality and completeness of the overall test definition
(test system, protocol, test acceptance criteria etc.)
3.2 Quality of the background provided concerning the purpose of the test method
3.3 Quality of the documentation and completeness of (a) standardised protocols (SOPs) and (b) prediction models

4. Data quality

4.1 Overall quality of the evaluated data
4.2 Sufficiency of the evaluated data in view of the study objective
4.3 Quality of the reference data for evaluating reliability and relevance[5]

5. Test materials

5.1 Sufficiency of the number of evaluated test items in view of the study objective
5.2 Representativeness of the test items with respect to the applicability domain

6. Within-laboratory reproducibility (Module 2)

6.1 Assessment of repeatability and reproducibility in the same laboratory
6.2 Conclusion on within-laboratory reproducibility as assessed by the study

[5] OECD guidance document Nr. 34 on validation defines relevance as follows: "Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It sis the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of accuracy (concordance) of a test method."

7. Transferability (Module 3)

7.1 Quality of design and analysis of the transfer phase

7.2 Conclusion on transferability to a second laboratory as assessed by the study

In particular: where critical issues that may impact on transferability identified or addressed?

8. Between-laboratory reproducibility (Module 4)

8.1 Assessment of reproducibility in different laboratories

8.2 Conclusion on reproducibility as assessed by the study

9. Predictive capacity (Module 5) N.B. Predictive capacity was outside the scope of the study

9.1 Adequacy of the assessment of the predictive capacity in view of the purpose

9.2 Overall relevance (biological relevance and accuracy) of the test method in view of the purpose

10. Applicability domain (Module 6) N.B. Since this study is not a full validation study, the assessment of the applicability domain is rather limited

10.1 Appropriateness of study design to conclude on applicability domain, limitations and exclusions

10.2 Quality of the description of applicability domain, limitations, exclusions

11. Performance standards (Module 7) N.B. Not applicable to the current study.

11.1 Adequacy of the proposed Essential Test Method Components

11.2 Adequacy of the Reference Chemicals

11.3. Adequacy of the defined Accuracy Values

12. Readiness for standardised use

12.1 Assessment of the readiness for regulatory purposes

12.2. Assessment of the readiness for other uses (in house screening etc.)

12.3 Critical aspects impacting on standardised use

12.4 Gap analysis

Identify, if appropriate, gaps in the study design and/or execution that impact on the stated study objective or the conclusions drawn.

13. Other considerations

Please address any other consideration you might have in relation to the proposed approach under this section.

14. Conclusions and recommendation

14.1 Summary of the study results and conclusions

14.2 Extent to which conclusions are justified by the study results alone

14.3 Extent to which conclusions are plausible in the context of existing information

14.4 Recommendations