

The Use of Scientifically-Validated *In Vitro* Tests for Embryotoxicity

At its 17th Meeting, held on 16-17 October 2001, the ECVAM Scientific Advisory Committee (ESAC) endorsed three *in vitro* methods for embryotoxicity testing as scientifically validated and ready for consideration for regulatory acceptance and application.

However, the ESAC recognised that these methods do not represent replacements for current animal tests for reproductive toxicity as a whole, although they could provide suitable means for reducing and/or refining the use of animal procedures in the context of a testing strategy or testing strategies.

The Committee also recommended that an ECVAM workshop should be held, to discuss with experts with relevant and complementary experience, the drafting of a guidance document on the applicability of the three scientifically-validated embryotoxicity test methods in the context of reproductive toxicity testing as a whole.

Michael Balls
Head of Unit
ECVAM
Institute for Health & Consumer Protection
Joint Research Centre
European Commission
Ispra, Italy

3 June 2002

Statement on the Scientific Validity of the Embryonic Stem Cell Test (EST) - an *In Vitro* Test for Embryotoxicity

At its 17th meeting, held on 16-17 October 2001 at the European Centre for the Validation of Alternative Methods (ECVAM), Ispra, Italy, the ECVAM Scientific Advisory Committee (ESAC)¹ unanimously endorsed the following statement:

“The results obtained with the embryonic stem cell test (EST)^{2,3} in the blind trial definitive phase of the ECVAM international validation study on *in vitro* tests for embryotoxicity were highly reproducible, both within and among the four laboratories that performed the test. The correlation between the *in vitro* data and *in vivo* data was good (accuracy 78%) according to the performance criteria defined. The test proved applicable to testing a diverse group of chemicals of different embryotoxic potentials (*non-embryotoxic*, *weakly embryotoxic* and *strongly embryotoxic*). The predictivity (100%) for *strongly embryotoxic* chemicals was excellent and the precision (81%) was good. The predictivity for *non* (72%) and *weak* (70%) embryotoxicants and the precision for *non-embryotoxic* compounds (70%) were sufficiently high ($\geq 65\%$).

The Committee therefore agrees with the conclusion from this formal validation study that the embryonic stem cell test (EST) is a scientifically validated test which is ready to be considered for regulatory purposes.”

The ESAC has been regularly kept informed of the progress of the study, and this endorsement is based on an assessment of various documents, including publications on the results of the prevalidation phase of the study³⁻⁵ and development of the prediction model,⁶⁻⁷ and in particular the final report on the performance of the EST in a multilaboratory blind trial on 20 coded chemicals, which is published in *ATLA*.^{8,9}

This study was conducted according to ECVAM's principles and procedures of validation,¹⁰⁻¹² and according to criteria laid down by the Organisation for Economic Cooperation and Development (OECD)¹³ and the US validation centre, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) at the National Institute of Environmental Health Sciences (NIEHS).¹⁴

Michael Balls
Head of Unit
ECVAM
Institute for Health & Consumer Protection
Joint Research Centre
European Commission
Ispra
Italy

Eva Hellsten
Head of Unit E.2
Environment Directorate General
European Commission
Brussels
Belgium

1 May 2002

1. The ESAC was established by the European Commission, and is composed of representatives of the EU Members States, industry, academia and animal welfare, together with representatives of the relevant Commission services. The following members of the ESAC were present at the meeting on 16-17 October 2001:

Dr Bas Blaauboer (ERGATT)

Mr Michael Balls (ECVAM -
Chairman)

Dr Philip Botham (ECETOC)

Ms Susanna Louhimies (DG ENV)

Professor José Castell (Spain)

Ms Beatrice Lucaroni (DG RTD)

Dr Bernward Garthoff (EFPIA)

Mr Lars Nørgaard (DG ENTR)

Professor André Guillouzo (France)

Mr Juan Riego Sintes (ECB)

Dr Coenraad Hendriksen (The Netherlands)

Mr Enrico Sabbioni (ECVAM)

Dr Maggy Jennings (EUROGROUP for Animal
Welfare)

Mr Andrew Worth (ECVAM -
Secretary)

Professor Vera Rogiers (Belgium)

Dr Odile de Silva (COLIPA)

Professor Horst Spielmann (Germany)

Professor Helmut Tritthart (Austria)

Dr Matti Viluksela (Finland)

Professor Erik Walum (Sweden)

2. Spielmann, H., Pohl, I., Döring, B., Liebsch, M. & Moldenhauer, F. (1997). The embryonic stem cell test (EST), an *in vitro* embryotoxicity test using two permanent mouse cell lines: 3T3 fibroblasts and embryonic stem cells. *In Vitro Toxicology* **10**, 119-127.
3. Scholz, G., Pohl, I., Seiler, A., Bremer, S., Brown, N.A., Piersma, A.H., Holzhütter, G.H. & Spielmann, H. (1998). Ergebnisse der ersten Phase des ECVAM-Projektes zur Prävalidierung und Validierung von drei *in vitro* Embryotoxizitätstests. *ALTEX* **15**, 3-8.
4. Scholz, G., Genschow, E., Pohl, I., Bremer, S., Paparella, M., Raabe, H., Moldenhauer, F., Southee, J. & Spielmann, H. (1999). Prevalidation of the embryonic stem cell test (EST) - A new *in vitro* embryotoxicity test. *Toxicology in Vitro* **13**, 675-681.
5. Scholz, G., Pohl, I., Genschow, E., Klemm, M. & Spielmann, H. (1999). Embryotoxicity screening using ES cells *in vitro*: correlation to *in vivo* teratogenicity. *Cells Tissues Organs* **165**, 203-211.
6. Genschow, E., Scholz, G., Brown, N.A., Piersma, A.H., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S. & Spielmann, H. (1999). Die Entwicklung von Prädiktionsmodellen für drei *in vitro* Embryotoxizitätstests im Rahmen einer ECVAM Validierungsstudie. *ALTEX* **16**, 73-83.
7. Genschow, E., Scholz, G., Brown, N.A., Piersma, A.H., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S. & Spielmann, H. (2000). Development of prediction models for three *in vitro* embryotoxicity tests in an ECVAM Validation Study. *In Vitro and Molecular Toxicology* **13**, 51-65.

8. Genschow, G., Spielmann, H., Scholz, G., Seiler, A., Brown, N., Piersma, A., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S. & Becker, K. (2002). The ECVAM international validation study on *in vitro* embryotoxicity tests. Results of the definitive phase and evaluation of prediction models. *ATLA* **30**, 151-176.
9. Brown NA. (2002). Selection of test chemicals for the ECVAM international validation study on *in vitro* embryotoxicity tests. *ATLA* **30**, 177-198.
10. Balls, M., Blaauboer, B., Brusik, D., Frazier, J., Lamb, D., Pemberton, M., Reinhardt, C., Roberfroid, M., Rosenkranz, H., Schmid, B., Spielmann, H., Stamatii, A.L. & Walum, E. (1990). Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures. *ATLA* **18**, 313-337.
11. Balls, M., Blaauboer, B. J., Fentem, J.H., Bruner, L., Combes, R.D., Ekwall, B., Fielder, R.J., Guillouzo, A., Lewis, R.W., Lovell, D.P., Reinhardt, C.A., Repetto, G., Sladowski, D., Spielmann, H. & Zucco, F. (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations ECVAM workshop 5. *ATLA* **23**, 129-147
12. Curren, R.D., Southee J.A., Spielmann, H., Liebsch, M., Fentem, J.H. & Balls, M. (1995). The role of prevalidation in the development, validation and acceptance of alternative methods. The first report of the ECVAM task force on prevalidation. *ATLA* **23**, 211-217.
13. OECD (1996). Final Report of the OECD Workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Test Methods, 60pp. Paris, France: OECD.
14. NIEHS (1997). Validation and Regulatory Acceptance of Toxicological Test Methods. A Report of the Ad Hoc Interagency Coordinating Committee on the Validation of Alternative Methods, 105pp. Research Triangle Park, NC: NIEHS.

General information about the study

- A. In December 1996, following an open call-for-tender, an ECVAM contract was awarded to ZEBET / BgVV (Berlin, Germany), to plan and coordinate the prevalidation and validation of three embryotoxicity tests: a) a test employing embryonic stem cell lines; b) the micromass test; and c) the postimplantation rat whole-embryo culture assay. Phase I (1997) of the study³ was designed as a prevalidation phase, for test protocol optimisation and establishment of a comprehensive database of *in vivo* and *in vitro* data on embryotoxic compounds, for use in the prevalidation and validation of the three assays. Phase II (1998-2000) involved a formal validation trial⁸ conducted under blind conditions on 20 carefully selected test chemicals,⁹ which were coded and distributed to the participating laboratories. The results obtained were submitted to an independent statistician for analysis. Data analysis and preparation of the final report took

place during 1999-2000. The final report of the validation study was submitted on 10 March 2000 and was accepted by ECVAM as stated in a letter of 24 May 2000.

- B. The study was managed by a Management Team consisting of representatives of the ECVAM Task Force on Developmental Toxicology, under the chairmanship of Professor Horst Spielmann (ZEBET / BgVV, Centre for Documentation and Evaluation of Alternative Methods to Animal Experiments at the BgVV, Berlin, D). The following laboratories participated in the blind trial on the EST: ZEBET (the lead laboratory), ECVAM (Ispra, I), Novartis (Basel, CH) and Schering (Berlin, D).
- C. In the embryonic stem cell test (EST) two permanent mouse cell lines are used to assess the embryotoxic potential of test chemicals²⁻⁷: 3T3 fibroblasts and the embryonic stem cell line D3. Inhibition of differentiation and growth are determined in embryonic stem (ES) cells and compared to inhibition of growth in 3T3 fibroblasts, which serve as surrogate for adult cells. Three endpoints are used to classify the embryotoxic potential of test chemicals in the EST²⁻⁷: the inhibition of growth of ES cells and 3T3 fibroblasts (cytotoxicity measured in the MTT assay) by 50% of the control (IC₅₀D3, IC₅₀3T3) and the inhibition of the differentiation of ES cells into spontaneously contracting cardiomyocytes by 50% (ID₅₀).
- D. A prediction model was developed with the test compounds of the prevalidation study and evaluated using the six test chemicals of the preliminary phase of the validation study.²⁻⁷ As a next step, this model was applied to classify the embryotoxic potentials of the 14 test chemicals of the definitive phase of the validation study into three classes of embryotoxicity (*non*, *weak*, *strong*) on the basis of the *in vitro* data obtained in the four laboratories.^{8,9}

Comparing these *in vitro* classifications to the *in vivo* classifications independently assigned to the chemicals prior to the blind trial⁸, the following overall contingency statistics were obtained for the EST:⁷

	EST
Predictivity for non-embryotoxic	72%
Predictivity for weakly embryotoxic	70%
Predictivity for strongly embryotoxic	100%
Precision for non-embryotoxic	70%
Precision for weakly embryotoxic	83%
Precision for strongly embryotoxic	81%
Accuracy	78%

- E. In order to compare the results of the three embryotoxicity tests the overall contingency statistics are outlined in the following table.

	MM test	WEC	
		PM1	PM2
Predictivity for non-embryotoxic	57%	56%	70%

Predictivity for weakly embryotoxic	71%	75%	76%
Predictivity for strongly embryotoxic	100%	79%	100%
Precision for non-embryotoxic	80%	70%	80%
Precision for weakly embryotoxic	60%	45%	65%
Precision for strongly embryotoxic	69%	94%	100%
Accuracy	70%	68%	80%

F. The design of 3x3 contingency tables that were used in the statistical evaluation and the biostatistical performance criteria are shown in the following tables.

3x3 contingency table

<i>In vivo</i> class	Prediction		
	class 1	class 2	class 3
Not embryotoxic ¹	a	b	c
Weak embryotoxic ²	d	e	f
Strong embryotoxic ³	g	h	i

1) class 1; 2) class 2; 3) class 3

$$n = a+b+c+d+e+f+g+h+i$$

Precision for strong (respectively weak) embryotoxic chemicals is defined as the proportion of correctly classified strong (respectively weak) embryotoxic chemicals from the *in vitro* test versus chemicals that are strong (respectively weak) embryotoxic *in vivo*.⁶ Precision for non embryotoxic chemicals describes the proportion of the correctly predicted *non embryotoxic* chemicals relative to the non embryotoxic chemicals *in vivo*.

Statistics of 3x3 contingency table

Predictivity for non-embryotoxic chemicals	$\left(\frac{a}{a+d+g}\right) * 100$
Predictivity for weakly embryotoxic chemicals	$\left(\frac{e}{b+e+h}\right) * 100$
Predictivity for strongly embryotoxic chemicals	$\left(\frac{i}{c+f+i}\right) * 100$
Precision for non-embryotoxic chemicals	$\left(\frac{a}{a+b+c}\right) * 100$
Precision for weakly embryotoxic chemicals	$\left(\frac{e}{d+e+f}\right) * 100$
Precision for strongly embryotoxic chemicals	$\left(\frac{i}{g+h+i}\right) * 100$
Accuracy	$\left(\frac{a+e+i}{n}\right) * 100$

Predictivity for strong (respectively weak) embryotoxic chemicals is an estimate of the likelihood that a positive prediction in the test correctly identifies a strong (respectively weak) embryotoxic test chemical under the proposed condition of use. Predictivity for non embryotoxic chemicals is, therefore, the estimate of the likelihood that a prediction for non embryotoxic chemicals in the test correctly identifies a non embryotoxic test chemical. The definitions (*predictivity*, *precision*) used in the 3x3 contingency tables for *in vitro* toxicology tests were adapted from the definitions (*specificity*, *sensitivity*) of 2x2 contingency tables.⁷

G. The management team (MT) of the ECVAM validation study agreed on the following classification criteria:⁸

Since the test chemicals are assigned to three classes of embryotoxicity,^{7,8} 33% correct classifications can be expected purely by chance. In contrast, if two classes are used for classification, an a priori probability of 50% can be expected by chance. The criteria used by the MT of the present study to evaluate the performance of the tests are shown in the following table.

	Test performance (%)
By chance	33
Sufficient	≥ 65
Good	≥ 75
Excellent	≥ 85

This evaluation takes into account the inherent variability of the *in vivo* data.⁸ Thus an excellent performance was defined by a test performance of 85% for each of the performance criteria (accuracy, predictivity, precision), while the result was considered insufficient if the performance was below 65%.

Statement on the Scientific Validity of the Micromass Test - an *In Vitro* Test for Embryotoxicity

At its 17th meeting, held on 16-17 October 2001 at the European Centre for the Validation of Alternative Methods (ECVAM), Ispra, Italy, the ECVAM Scientific Advisory Committee (ESAC)¹ unanimously endorsed the following statement:

“The results obtained with the micromass test² in the blind trial definitive phase of the ECVAM international validation study on *in vitro* tests for embryotoxicity were highly reproducible, both within and among the four laboratories that performed the test. The correlation between the *in vitro* data and *in vivo* data was sufficient (accuracy 70%) according to the performance criteria defined. The test proved applicable to testing a diverse group of chemicals of different embryotoxic potentials (*non-embryotoxic*, *weakly embryotoxic*, and *strongly embryotoxic*). The predictivity (100%) for strong embryotoxic chemicals was excellent while the precision (69%) was sufficient. The predictivity for *non-embryotoxic* test chemicals (57%) and the precision for *weakly embryotoxic* test chemicals (60%) were insufficient (< 65%).

The Committee therefore agrees with the conclusion from this formal validation study that the micromass test is a scientifically validated test for identifying strong embryotoxic chemicals, which is ready to be considered for regulatory purposes.”

The ESAC has been regularly kept informed of the progress of the study, and this endorsement is based on an assessment of various documents, including publications on the results of the prevalidation phase of the study³ and development of the prediction model,⁴⁻⁵ and in particular the final report on the performance of the EST in a multilaboratory blind trial on 20 coded chemicals, which is published in *ATLA*.^{6, 7}

This study was conducted according to ECVAM's principles and procedures of validation,⁸⁻¹⁰ and according to criteria laid down by the Organisation for Economic Cooperation and Development (OECD)¹¹ and the US validation centre, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) at the National Institute of Environmental Health Sciences (NIEHS).¹²

Michael Balls
Head of Unit
ECVAM
Institute for Health & Consumer Protection
Joint Research Centre
European Commission
Ispra
Italy

Eva Hellsten
Head of Unit E.2
Environment Directorate General
European Commission
Brussels
Belgium

1 May 2002

1. The ESAC was established by the European Commission, and is composed of representatives of the EU Members States, industry, academia and animal welfare, together with representatives of the relevant Commission services. The following members of the ESAC were present at the meeting on 16-17 October 2001:

Dr Bas Blaauboer (ERGATT)

Dr Philip Botham (ECETOC)

Professor José Castell (Spain)

Dr Bernward Garthoff (EFPIA)

Professor André Guillouzo (France)

Dr Coenraad Hendriksen (The Netherlands)

Dr Maggy Jennings (EUROGROUP for Animal Welfare)

Professor Vera Rogiers (Belgium)

Dr Odile de Silva (COLIPA)

Professor Horst Spielmann (Germany)

Professor Helmut Tritthart (Austria)

Dr Matti Viluksela (Finland)

Professor Erik Walum (Sweden)

Mr Michael Balls (ECVAM - Chairman)

Ms Susanna Louhimies (DG ENV)

Ms Beatrice Lucaroni (DG RTD)

Mr Lars Nørgaard (DG ENTR)

Mr Juan Riego Sintes (ECB)

Mr Enrico Sabbioni (ECVAM)

Mr Andrew Worth (ECVAM - Secretary)

2. Flint, O.P. & Orton, T.C. (1983). An *in vitro* assay for teratogens with cultures of rat embryonic midbrain and limb cells. *Toxicology and Applied Pharmacology* **76**, 383-395.
3. Scholz, G., Pohl, I., Seiler, A., Bremer, S., Brown, N.A., Piersma, A.H., Holzhütter, G.H. & Spielmann, H. (1998). Ergebnisse der ersten Phase des ECVAM-Projektes zur Prävalidierung und Validierung von drei *in vitro* Embryotoxizitätstests. *ALTEX* **15**, 3-8.
4. Genschow, E., Scholz, G., Brown, N.A., Piersma, A.H., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S. & Spielmann, H. (1999). Die Entwicklung von Prädiktionsmodellen für drei *in vitro* Embryotoxizitätstests im Rahmen einer ECVAM Validierungsstudie. *ALTEX* **16**, 73-83.
5. Genschow, E., Scholz, G., Brown, N.A., Piersma, A.H., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S. & Spielmann, H. (2000). Development of prediction models for three *in vitro* embryotoxicity tests in an ECVAM Validation Study. *In Vitro and Molecular Toxicology* **13**, 51-65.
6. Genschow, G., Spielmann, H., Scholz, G., Seiler, A., Brown, N., Piersma, A., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S. & Becker, K. (2002). The ECVAM international validation study on *in vitro* embryotoxicity tests. Results of the definitive phase and evaluation of prediction models. *ATLA* **30**, 151-176.
7. Brown, N. (2002). Selection of test chemicals for the ECVAM international validation study on *in vitro* embryotoxicity tests. *ATLA* **30**, 177-198.

8. Balls, M., Blaauboer, B., Brusik, D., Frazier, J., Lamb, D., Pemberton, M., Reinhardt, C., Roberfroid, M., Rosenkranz, H., Schmid, B., Spielmann, H., Stamatii, A.L. & Walum, E. (1990). Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures. *ATLA* **18**, 313-337.
9. Balls, M., Blaauboer, B. J., Fentem, J.H., Bruner, L., Combes, R.D., Ekwall, B., Fielder, R.J., Guillouzo, A., Lewis, R.W., Lovell, D.P., Reinhardt, C.A., Repetto, G., Sladowski, D., Spielmann, H. & Zucco, F. (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations ECVAM workshop 5. *ATLA* **23**, 129-147
10. Curren, R.D., Southee J.A., Spielmann, H., Liebsch, M., Fentem, J.H. & Balls, M. (1995). The role of prevalidation in the development, validation and acceptance of alternative methods. The first report of the ECVAM task force on prevalidation. *ATLA* **23**, 211-217.
11. OECD (1996). Final Report of the OECD Workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Test Methods, 60pp. Paris: OECD.
12. NIEHS (1997). Validation and Regulatory Acceptance of Toxicological Test Methods. A Report of the Ad Hoc Interagency Coordinating Committee on the Validation of Alternative Methods, 105pp. Research Triangle Park, NC: NIEHS.

General information about the study

- A. In December 1996, following an open call-for-tender, an ECVAM contract was awarded to ZEBET / BgVV (Berlin, Germany), to plan and coordinate the prevalidation and validation of three embryotoxicity tests: a) a test employing embryonic stem cell lines; b) the micromass test; and c) the postimplantation rat whole-embryo culture assay. Phase I (1997) of the study³ was designed as a prevalidation phase, for test protocol optimisation and establishment of a comprehensive database of *in vivo* and *in vitro* data on embryotoxic compounds, for use in the prevalidation and validation of the three assays. Phase II (1998-2000) involved a formal validation trial⁵ conducted under blind conditions on 20 carefully selected test chemicals,⁶ which were coded and distributed to the participating laboratories. The results obtained were submitted to an independent statistician for analysis. Data analysis and preparation of the final report took place during 1999-2000. The final report of the validation study was submitted on 10 March, 2000 and was accepted by ECVAM as stated in a letter of 24 May 2000.
- B. The study was managed by a Management Team consisting of representatives of the ECVAM Task Force on Developmental Toxicology, under the chairmanship of Professor Horst Spielmann (ZEBET/BgVV, Centre for Documentation and Evaluation of Alternative Methods to Animal Experiments at the BgVV, Berlin, Germany). The following laboratories participated in the blind trial on the micromass test: St. George's Hospital Medical School (SGHMS, University of

London, UK, the lead laboratory), National Institute of Public Health and the Environment (RIVM, Bilthoven, The Netherlands), KTL Finland (Kuopio, Finland) and Synthelabo Recherche (Gargenville, France).

- C. In the micromass test aggregates of undifferentiated mesenchyme cells of early rat embryo limbs are cultured in small volumes at high density. They will form numerous small foci of differentiating chondrocytes within a background of undifferentiated cells.² In principle the micromass test is based on the ability of test chemicals to inhibit the formation of foci. Two endpoints are determined in the micromass test². In the present study, differentiation was assessed by alcian blue staining of cartilage and growth of micromass cells was determined by neutral red uptake.³⁻⁶ Since biostatistical evaluation of the two endpoints proved that both of them provided the same information,⁵ in the definitive phase of the study, 50% inhibition of differentiation assessed by alcian blue staining (ID₅₀) was used as the only endpoint to classify the embryotoxic potential of test chemicals.
- D. A prediction model was developed using the six test chemicals of the preliminary phase of the validation study.⁵ As a next step, this model was applied to classify the embryotoxic potentials of the 14 test chemicals of the definitive phase of the validation study into three classes of embryotoxicity (*non*, *weak*, *strong*) on the basis of the *in vitro* data obtained in the four laboratories.^{6,7}

Comparing the *in vitro* classifications with the *in vivo* classifications independently assigned to the chemicals prior to the blind trial,⁷ the following overall contingency statistics were obtained for the micromass test⁶:

	MM test
Predictivity for non-embryotoxic	57%
Predictivity for weakly embryotoxic	71%
Predictivity for strongly embryotoxic	100%
Precision for non-embryotoxic	80%
Precision for weakly embryotoxic	60%
Precision for strongly embryotoxic	69%
Accuracy	70%

- E. In order to compare the results of the three embryotoxicity tests, the overall contingency statistics are outlined in the following table.

	PM1	WEC PM2	EST
Predictivity for non-embryotoxic	56%	70%	72%
Predictivity for weakly embryotoxic	75%	76%	70%
Predictivity for strongly embryotoxic	79%	100%	100%

Precision for non-embryotoxic	70%	80%	70%
Precision for weakly embryotoxic	45%	65%	83%
Precision for strongly embryotoxic	94%	100%	81%
Accuracy	68%	80%	78%

- F. The design of 3x3 contingency tables that were used in the statistical evaluation and the biostatistical performance criteria are shown in the following tables.

3x3 contingency table

<i>In vivo</i> class	Prediction		
	class 1	class 2	class 3
Not embryotoxic ¹	a	b	c
Weak embryotoxic ²	d	e	f
Strong embryotoxic ³	g	h	i

1) class 1; 2) class 2; 3) class 3

$$n = a+b+c+d+e+f+g+h+i$$

Precision for strong (respectively weak) embryotoxic chemicals is defined as the proportion of correctly classified strong (respectively weak) embryotoxic chemicals from the *in vitro* test versus chemicals that are strong (respectively weak) embryotoxic *in vivo*.⁷ Precision for non embryotoxic chemicals describes the proportion of the correctly predicted *non embryotoxic* chemicals relative to the non embryotoxic chemicals *in vivo*.

Statistics of 3x3 contingency table

Predictivity for non-embryotoxic chemicals	$\left(\frac{a}{a+d+g}\right) * 100$
Predictivity for weakly embryotoxic chemicals	$\left(\frac{e}{b+e+h}\right) * 100$
Predictivity for strongly embryotoxic chemicals	$\left(\frac{i}{c+f+i}\right) * 100$
Precision for non-embryotoxic chemicals	$\left(\frac{a}{a+b+c}\right) * 100$
Precision for weakly embryotoxic chemicals	$\left(\frac{e}{d+e+f}\right) * 100$
Precision for strongly embryotoxic chemicals	$\left(\frac{i}{g+h+i}\right) * 100$
Accuracy	$\left(\frac{a+e+i}{n}\right) * 100$

Predictivity for strong (respectively weak) embryotoxic chemicals is an estimate of the likelihood that a positive prediction in the test correctly identifies a strong (respectively weak) embryotoxic test chemical under the proposed condition of use. Predictivity for non embryotoxic chemicals is, therefore, the estimate of the likelihood that a prediction for non embryotoxic chemicals in the test correctly identifies a non embryotoxic test chemical. The definitions (*predictivity*, precision) used in the 3x3 contingency tables for *in vitro* toxicology tests were adapted from the definitions (*specificity*, sensitivity) of 2x2 contingency tables.⁸

- G. The management team (MT) of the ECVAM validation study agreed on the following classification criteria⁶:

Since the test chemicals are assigned to three classes of embryotoxicity,^{5,6} 33% of correct classifications can be expected purely by chance. In contrast, if two classes are used for classification, an a priori probability of 50% can be expected by chance. The criteria used by the MT of the present study to evaluate the performance of the tests are shown in the following table.

	Test performance (%)
By chance	33
Sufficient	≥ 65
Good	≥ 75
Excellent	≥ 85

This evaluation takes into account the inherent variability of the *in vivo* data.⁶ Thus an excellent performance was defined by a test performance of 85% for each of the performance criteria (accuracy, predictivity, precision), while the result was considered insufficient, if the performance was below 65%.

Statement on the Scientific Validity of the Postimplantation Rat Whole-Embryo Culture Assay -an *In Vitro* Test for Embryotoxicity

At its 17th meeting, held on 16-17 October 2001 at the European Centre for the Validation of Alternative Methods (ECVAM), Ispra, Italy, the ECVAM Scientific Advisory Committee (ESAC)¹ unanimously endorsed the following statement:

“The results obtained with the postimplantation rat whole-embryo culture (WEC) assay² in the blind trial definitive phase of the ECVAM international validation study on *in vitro* tests for embryotoxicity were reproducible, both within and among the four laboratories that performed the test. The correlation between the *in vitro* data and *in vivo* data was good (accuracy 80%) according to the performance criteria defined, when cytotoxicity data provided by the cytotoxicity test with 3T3 cells obtained in the embryonic stem cell test (EST) were included (prediction model PM2). The test proved applicable to testing a diverse group of chemicals of different embryotoxic potentials (*non-embryotoxic*, *weakly embryotoxic*, and *strongly embryotoxic*). The predictivity and precision for *strongly embryotoxic* test chemicals were excellent (100%). The predictivity for *non-embryotoxic* compounds (70%) and the precision for *weakly embryotoxic* compounds (65%) were sufficiently high ($\geq 65\%$).

The Committee therefore agrees with the conclusion from this formal validation study that the postimplantation rat whole-embryo culture assay is a scientifically validated test which is ready to be considered for regulatory purposes.”

The ESAC has been regularly kept informed of the progress of the study, and this endorsement is based on an assessment of various documents, including publications on the results of the prevalidation phase of the study³ and development of the prediction model,^{4,5} and in particular the final report on the performance of the micromass test in a multi-laboratory blind trial on 20 coded chemicals, which is published in *ATLA*.^{6,7}

This study was conducted according to ECVAM's principles and procedures of validation,⁸⁻¹⁰ and according to criteria laid down by the Organisation for Economic Cooperation and Development (OECD)¹¹ and the US validation centre, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) at the National Institute of Environmental Health Sciences (NIEHS).¹²

Michael Balls
Head of Unit
ECVAM
Institute for Health & Consumer Protection
Joint Research Centre
European Commission
Ispra
Italy

Eva Hellsten
Head of Unit E.2
Environment Directorate General
European Commission
Brussels
Belgium

1 May 2002

1. The ESAC is composed of representatives of the EU Members States, industry, academia and animal welfare, together with representatives of the relevant Commission services. The following members of the ESAC were present at the meeting on 16-17 October 2001:

Dr Bas Blaauboer (ERGATT)

Dr Philip Botham (ECETOC)

Professor José Castell (Spain)

Dr Bernward Garthoff (EFPIA)

Professor André Guillouzo (France)

Dr Coenraad Hendriksen (The Netherlands)

Dr Maggy Jennings (EUROGROUP for Animal Welfare)

Professor Vera Rogiers (Belgium)

Dr Odile de Silva (COLIPA)

Professor Horst Spielmann (Germany)

Professor Helmut Tritthart (Austria)

Dr Matti Viluksela (Finland)

Professor Erik Walum (Sweden)

Mr Michael Balls (ECVAM - Chairman)

Ms Susanna Louhimies (DG ENV)

Ms Beatrice Lucaroni (DG RTD)

Mr Lars Nørgaard (DG ENTR)

Mr Juan Riego Sintes (ECB)

Mr Enrico Sabbioni (ECVAM)

Mr Andrew Worth (ECVAM - Secretary)

2. Brown, N.A. & Fabro, S. (1981). Quantitation of rat embryo development *in vitro*: a morphological scoring system. *Teratology* **24**, 65-78.
3. Scholz, G., Pohl, I., Seiler, A., Bremer, S., Brown, N.A., Piersma, A.H., Holzhütter, G.H. & Spielmann, H. (1998). Ergebnisse der ersten Phase des ECVAM-Projektes zur Prävalidierung und Validierung von drei *in vitro* Embryotoxizitätstests. *ALTEX* **15**, 3-8.
4. Genschow, E., Scholz, G., Brown, N.A., Piersma, A.H., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S. & Spielmann, H. (1999). Die Entwicklung von Prädiktionsmodellen für drei *in vitro* Embryotoxizitätstests im Rahmen einer ECVAM Validierungsstudie. *ALTEX* **16**, 73-83.
5. Genschow, E., Scholz, G., Brown, N.A., Piersma, A.H., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S. & Spielmann, H. (2000). Development of prediction models for three *in vitro* embryotoxicity tests in an ECVAM Validation Study. *In Vitro and Molecular Toxicology* **13**, 51-65.
6. Genschow, G., Spielmann, H., Scholz, G., Seiler, A., Brown, N., Piersma, A., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S. & Becker, K. (2002). The ECVAM international validation study on *in vitro* embryotoxicity tests. – Results of the definitive phase and evaluation of prediction models. *ATLA* **30**, 151-176.
7. Brown, N. (2002). Selection of test chemicals for the ECVAM international validation study on *in vitro* embryotoxicity tests. *ATLA* **30**, 177-198.

8. Balls, M., Blaauboer, B., Brusik, D., Frazier, J., Lamb, D., Pemberton, M., Reinhardt, C., Roberfroid, M., Rosenkranz, H., Schmid, B., Spielmann, H., Stamatii, A.L. & Walum, E. (1990). Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures. *ATLA* **18**, 313-337.
9. Balls, M., Blaauboer, B. J., Fentem, J.H., Bruner, L., Combes, R.D., Ekwall, B., Fielder, R.J., Guillouzo, A., Lewis, R.W., Lovell, D.P., Reinhardt, C.A., Repetto, G., Sladowski, D., Spielmann, H. & Zucco, F. (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations ECVAM workshop 5. *ATLA* **23**, 129-147
10. Curren, R.D., Southee J.A., Spielmann, H., Liebsch, M., Fentem, J.H. & Balls, M. (1995). The role of prevalidation in the development, validation and acceptance of alternative methods. The first report of the ECVAM task force on prevalidation. *ATLA* **23**, 211-217.
11. OECD (1996). Final Report of the OECD Workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Test Methods, 60pp. Paris, France: OECD.
12. NIEHS (1997). Validation and Regulatory Acceptance of Toxicological Test Methods. A Report of the Ad Hoc Interagency Coordinating Committee on the Validation of Alternative Methods, 105pp. Research Triangle Park, NC: NIEHS.

General information about the study

- A. In December 1996, following an open call-for-tender, an ECVAM contract was awarded to ZEBET/BgVV (Berlin, Germany), to plan and coordinate the prevalidation and validation of three embryotoxicity tests: a) a test employing embryonic stem cell lines; b) the micromass test; and c) the postimplantation rat whole-embryo culture assay. Phase I (1997) of the study³ was designed as a prevalidation phase, for test protocol optimisation and establishment of a comprehensive database of *in vivo* and *in vitro* data on embryotoxic compounds, for use in the prevalidation and validation of the three assays. Phase II (1998-2000) involved a formal validation trial⁶ conducted under blind conditions on 20 carefully selected test chemicals,⁷ which were coded and distributed to the participating laboratories. The results obtained were submitted to an independent statistician for analysis. Data analysis and preparation of the final report took place during 1999-2000. The final report of the validation study was submitted on 10 March 10 2000 and was accepted by ECVAM as stated in a letter of 24 May 2000.
- B. The study was managed by a Management Team consisting of representatives of the ECVAM Task Force on Developmental Toxicology, under the chairmanship of Professor Horst Spielmann (ZEBET/BgVV, Centre for Documentation and Evaluation of Alternative Methods to Animal Experiments at the BgVV, Berlin, Germany). The following laboratories participated in the blind trial on the postimplantation rat whole-embryo culture assay: National Institute of Public Health and the Environment (RIVM, Bilthoven, The Netherlands, the lead

laboratory) St. George's Hospital Medical School (SGHMS, University of London, UK), Novartis (Basel, Switzerland) and AstraZeneca CTL (Macclesfield, UK).

C. In the postimplantation rat whole-embryo culture assay an elaborate morphological scoring system² is used to assess signs of malformation or retardation in embryos exposed over a 48 hr period as well as embryonic death. Since biostatistical evaluation had shown during development of the prediction model that the scoring system could not reproducibly be applied even in experienced laboratories, the following endpoints were used to classify the embryotoxic potentials of test chemicals⁵:

Prediction model 1 (PM1):

- IC₅₀ for malformation (the concentration at which 50% of the embryos are malformed) and
- IC_{NOEC} for TMS (the maximum concentration that has no effect on the total morphological score (TMS)).

Prediction model 2 (PM2):

- IC_{max} for malformation (the lowest concentration that shows a maximum rate of malformation and
- the relative distance between IC₅₀ 3T3 (inhibition of growth of 3T3 cells by 50% of the control in the neutral red uptake test) and IC_{NOEC} for TMS.

D. Two versions of the prediction model (PM1 and PM2) were developed using the six test chemicals of the preliminary phase of the validation study.⁵ Because the PM1 of the postimplantation rat whole-embryo culture assay takes only parameters of differentiation and development into account, and not cytotoxicity data, a second prediction model (PM2) was developed by incorporating cytotoxicity data for the differentiated mouse fibroblast cell line 3T3 (IC₅₀ 3T3) which were derived from the embryonic stem cell test (EST). As a next step, PM1 and PM2 were applied to classify the embryotoxic potentials of the 14 test chemicals of the definitive phase of the validation study into three classes of embryotoxicity (*non*, *weak*, *strong*) on the basis of the *in vitro* data obtained in the four laboratories.^{6,7} Comparing these *in vitro* classifications with the *in vivo* classifications independently assigned to the chemicals prior to the blind trial,⁷ the following overall contingency statistics were obtained for the postimplantation rat whole-embryo culture assay⁶:

	WEC	
	PM1	PM2
Predictivity for non-embryotoxic	56%	70%
Predictivity for weakly embryotoxic	75%	76%
Predictivity for strongly embryotoxic	79%	100%
Precision for non-embryotoxic	70%	80%
Precision for weakly embryotoxic	45%	65%
Precision for strongly embryotoxic	94%	100%

Accuracy 68% 80%

- E. In order to compare the results of the three embryotoxicity tests the overall contingency statistics are outlined in the following table.

	MM test	EST
Predictivity for non-embryotoxic	57%	72%
Predictivity for weakly embryotoxic	71%	70%
Predictivity for strongly embryotoxic	100%	100%
Precision for non-embryotoxic	80%	70%
Precision for weakly embryotoxic	60%	83%
Precision for strongly embryotoxic	69%	81%
Accuracy	70%	78%

- F. The design of 3x3 contingency tables that were used in the statistical evaluation and the biostatistical performance criteria are shown in the following tables.

3x3 contingency table

<i>In vivo</i> class	Prediction		
	class 1	class 2	class 3
Not embryotoxic ¹	a	b	c
Weak embryotoxic ²	d	e	f
Strong embryotoxic ³	g	h	i

1) class 1; 2) class 2; 3) class 3
 $n = a+b+c+d+e+f+g+h+i$

Precision for strongly (weak) embryotoxic chemicals is defined as the proportion of correctly classified strong (weakly) embryotoxic chemicals from the *in vitro* test versus chemicals that are strong (weakly) embryotoxic *in vivo*.⁷ Precision for non-embryotoxic chemicals describes the proportion of the correctly predicted *non-embryotoxic* chemicals relative to the non embryotoxic chemicals *in vivo*.

Statistics of 3x3 contingency table

Predictivity for non-embryotoxic chemicals	$\left(\frac{a}{a+d+g}\right) * 100$
Predictivity for weakly embryotoxic chemicals	$\left(\frac{e}{b+e+h}\right) * 100$
Predictivity for strongly embryotoxic chemicals	$\left(\frac{i}{c+f+i}\right) * 100$

Precision for non-embryotoxic chemicals	$\left(\frac{a}{a+b+c}\right)*100$
Precision for weakly embryotoxic chemicals	$\left(\frac{e}{d+e+f}\right)*100$
Precision for strongly embryotoxic chemicals	$\left(\frac{i}{g+h+i}\right)*100$
Accuracy	$\left(\frac{a+e+i}{n}\right)*100$

Predictivity for strong (respectively weak) embryotoxic chemicals is an estimate of the likelihood that a positive prediction in the test correctly identifies a strong (respectively weak) embryotoxic test chemical under the proposed condition of use. Predictivity for non embryotoxic chemicals is, therefore, the estimate of the likelihood that a prediction for non embryotoxic chemicals in the test correctly identifies a non embryotoxic test chemical. The definitions (*predictivity*, *precision*) used in the 3x3 contingency tables for *in vitro* toxicology tests were adapted from the definitions (*specificity*, *sensitivity*) of 2x2 contingency tables.⁸

- G. The management team (MT) of the ECVAM validation study agreed on the following classification criteria⁶:

Since the test chemicals are assigned to three classes of embryotoxicity,^{6,7} 33% of correct classifications can be expected purely by chance. In contrast, if two classes are used for classification, an a priori probability of 50% can be expected by chance. The criteria used by the MT of the present study to evaluate the performance of the tests are shown in the following table.

	Test performance (%)
By chance	33
Sufficient	≥ 65
Good	≥ 75
Excellent	≥ 85

This evaluation takes into account the inherent variability of the *in vivo* data.⁷ Thus an excellent performance was defined by a test performance of 85% for each of the performance criteria (accuracy, predictivity, precision), whereas the result was considered insufficient, if the performance was below 65%.