Screening Chemicals for Reproductive Toxicity: the Current Alternatives

The Report and Recommendations of an ECVAM/ETS Workshop (ECVAM Workshop 12)\(^1,2\)

Nigel A. Brown,\(^3\) Horst Spielmann,\(^4\) Rudolf Bechter,\(^5\) Oliver P. Flint,\(^6\) Stuart J. Freeman,\(^7\) Richard J. Jelínek,\(^8\) Elisabeth Koch,\(^5\) Heinz Nau,\(^9\) Derek R. Newall,\(^10\) Anthony K. Palmer,\(^11\) Jean-Yves Renault,\(^12\) Marina F. Repetto,\(^13\) Richard Vogel\(^14\) and Richard Wiger\(^15\)

\(^3\)MRC Experimental Embryology and Teratology Unit, St. George’s Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK; \(^4\)ZEBET, BgVV, Diedersdorfer Weg 1, 12277 Berlin, Germany; \(^5\)Sandoz Pharma Limited, Drug Safety Assessment, 4002 Basle, Switzerland; \(^6\)Bristol-Myers Squibb, Pharmaceutical Research Institute, 6000 Thompson Road, Syracuse, NY 13221, USA; \(^7\)SmithKline Beecham Consumer Healthcare, St George’s Avenue, Weybridge, Surrey KT13 0DE, UK; \(^8\)Department of Histology and Embryology, 3rd Medical Faculty, Charles University, Ruska 87, 100 00 Prague 10, Czech Republic; \(^9\)Freie Universität of Berlin, Institute for Toxicology and Embryopharmacology, Garsystrasse 5, 1000 Berlin 33, Germany; \(^10\)GlaxoWellcome Research and Development, Genetic and Reproductive Toxicology, Ware, Herts. SG12 0DP, UK; \(^11\)Huntingdon Research Centre, Department of Toxicology, Huntingdon, Cambs. PE18 6ES, UK; \(^12\)Rhône-Poulenc Rorer, Drug Safety NW 9, 500 Arcola Road, Collegeville, PA 19426-0107, USA; \(^13\)Rhône-Poulenc, Secteur Agro, 06903 Sophia Antipolis Cedex, France; \(^14\)BgVV, Thiellallee 88-92, 14191 Berlin, Germany; \(^15\)National Institute of Public Health, Department of Environmental Medicine, Geitmyrsveien 75, 0462 Oslo 4, Norway

Preface

This is the report of the twelfth of a series of workshops organised by the European Centre for the Validation of Alternative Methods (ECVAM). ECVAM’s main goal, as defined in 1993 by its Scientific Advisory Committee, is to promote the scientific and regulatory acceptance of alternative methods which are of importance to the biosciences and which reduce, refine or replace the use of laboratory animals. One of the first priorities set by ECVAM was the implementation of procedures which would enable it to become better informed about the state-of-the-art of non-animal test development and validation, and the potential for the possible incorporation of alternative tests into regulatory procedures. It was decided that this would be best achieved by the organisation of ECVAM workshops on specific topics, at which small groups of invited experts would review the current status of various types of in vitro tests and their potential uses, and make recommendations about the best ways forward (1).

The workshop on Screening Chemicals for Reproductive Toxicity was held in Angera,
Italy, on 22–26 February 1994, under the chairmanship of Nigel Brown (St George’s Hospital Medical School, London, UK), and was co-organised by Horst Spielmann (ZEBET, Berlin, Germany). The participants comprised scientists working for academic, industrial and regulatory organisations, from Europe and the USA. This particular ECVAM workshop was organised jointly with the European Teratology Society (ETS). The background to the workshop is outlined below.

Within the European Union (EU), a hierarchical decision-tree approach is used for testing new substances for reproductive toxicity (2). A positive test result may cause the chemical to be classified and labelled according to one of the categories (“toxic to reproduction”) defined in the seventh amendment to the Dangerous Substances Directive (Council Directive 67/548/EEC). For base-set chemicals (that is, those with a production of less than one tonne per annum or less than five tonnes in total), there is a theoretical requirement for screening for reproductive toxicity; however, as yet, there is no agreement on which screen(s) should be used. The Organisation for Economic Cooperation and Development (OECD) has proposed two draft guidelines for in vitro screening: draft guideline 421 (reproduction/developmental toxicity screening test [3]) and draft guideline 422 (combined repeated dose toxicity study with the reproduction/developmental toxicity screening test [4]). When the EU Competent Authorities considered these OECD draft guidelines in 1993, they could not reach agreement on their acceptability; a major reason for this was the lack of validated alternative methods for screening chemicals for reproductive toxicity.

Thus, the objective of this ECVAM/ETS workshop was to review the state-of-the-art of in vitro screens for reproductive toxicity. As a result of discussions which took place at the workshop, a number of recommendations have been made concerning in vivo and in vitro reproductive toxicity testing, which are outlined at the end of this report.

Introduction

There has been a great deal of research and debate during the past 15 years concerning the use of alternatives to living mammals for testing the potential reproductive toxicities of chemical and physical agents, and mixtures. This area has been reviewed on several occasions, most recently at an international conference held in Ottawa, Canada, in 1992 (5). The starting premises for this particular workshop, accepted by all of the participants, were that: a) the use of in vitro methods is well established and they are invaluable for conducting mechanistic reproductive toxicity studies; and b) in vitro methods already play a valuable role in so-called “secondary testing”, that is, in the screening of series of structurally related chemicals, when at least one of the chemicals is of known reproductive toxicity in vivo and it has been demonstrated that these effects can be mimicked in an alternative test system.

Thus, this workshop concentrated on ways to screen new substances (particularly, non-pharmaceuticals) for their potential adverse effects on reproduction, by using methods which could be incorporated into guidelines drawn up to satisfy mandatory notification requirements. A detailed critique of the two draft OECD test guidelines 421 and 422 was not the major purpose of the workshop; nevertheless, the draft guidelines were considered, so that the advantages and limitations of the in vitro screens could be compared with the proposed in vivo approaches. Throughout the workshop, the following questions were borne in mind: a) if screening was to become mandatory, how should positive and negative results be interpreted and used within the testing strategy?; b) should new substances be labelled solely on the basis of results obtained in screens?; c) what would be the real benefits of using alternative methods (for example, the earlier detection of probable hazards and/or elimination of the need for further [animal] testing for certain chemicals)?; d) which tests could be incorporated into a screening battery at the present time?; and e) what are the merits of in vitro screens compared with simplified in vivo screens?

Reproduction is a continuous cycle, but for the purposes of toxicity testing it is broadly divided into pregnancy in females (during which period pre-natal or post-natal developmental toxicity may be induced), and the remainder of the cycle in both males and females (during which period fertility may be impaired). The majority of research into the
development of alternative tests has concentrated specifically on teratogenicity, which is only one manifestation of embryotoxicity, which, in turn, is one aspect of developmental toxicity. This is, no doubt, because of the potential complexity of, and multiple targets for, adverse effects on fertility, which include effects on libido, sexual behaviour, spermatogenesis, oogenesis, hormonal activity or physiological response, which could interfere with the capacity to fertilise, fertilisation itself, or the development of the zygote up to and including implantation.

In Vivo Testing for Classification and Labelling Purposes

Due to the complexity of the reproductive cycle, from gamete maturation through to implantation of the early embryo into the maternal uterus, and because of the lack of validated alternative tests for most steps in the cycle, testing in living animals is the only option currently available for assessing the possible effects of chemicals on reproduction. The draft OECD testing guidelines 421 and 422 have been proposed for base-set testing of new chemicals on the basis that they will enable major effects, but not all potential hazards, to be detected. However, effects on male and female fertility cannot be evaluated separately in these tests. Therefore, it is recommended that one or more of the following procedures are included: for males, detailed histological examination of the testes and epididymides, semen analysis and/or flow cytometric analysis of spermatogenic cell types; for females, detailed histological examination of the ovary and vaginal cytology.

Fertility

Comprehensive multigeneration studies are currently undertaken to provide information about the effects of chemicals on all aspects of the highly complex reproductive cycle (6). For base-level testing of new substances, the workshop participants agreed that these studies were inappropriate, both because of the number of animals used, and because of the time and expense involved. Nevertheless, it was also agreed, in general, that in vitro screening will never be able to cover all of the aspects of "fertility", because some of these require integrated functions which are found only in living animals. Thus, the key question is whether sufficient information can be derived from alternative tests to be able to confidently classify and label a substance as toxic to the reproductive system, given the a priori limitations.

Female fertility

Alternative approaches

Some aspects of female reproductive function can be modelled in vitro, and several cellular components of the female reproductive organs can be maintained in culture (7). Although such systems have been used extensively in reproductive endocrinology studies and in other basic research investigations, none of them have been used or validated as toxicity screens. However, several of these systems may be useful for specialised toxicological studies, with appropriate method development and validation. It is possible that, in the long-term future, a battery of such systems would be able to cover a large proportion of the female reproductive cycle.

In females, the proliferation of primordial germ cells and the initial steps of meiosis occur long before birth. From puberty onwards, a small number of the primary oocytes complete oogenesis and are released from the ovary. Currently, there is intense interest in (and controversy about) methods for the culture and maturation of primordial germ cells, because of the need for donor oocytes in infertility treatment. When such methods have been perfected for laboratory animals, they may be of use in toxicological studies. Ovarian somatic cells (granulosa, thecal and stromal cells) can be maintained in culture (7), and any adverse effects can be assessed by examining cell morphology, and by determining cell viability and hormonal responsiveness.

Techniques for in vitro fertilisation (including functional maturation of spermatozoa) are routine, both clinically and in farm and laboratory animals. Although there are some examples of toxicological studies having been undertaken (8), the methods employed have not been validated for testing purposes. For example, the mammalian preimplantation period can be investigated by culturing embryos from the first cleavage divisions up to implantation, and toxicologi-
cal investigations at the chromosomal level have been published (9). Methods for determining parental imprinting during the preimplantation period should be developed, to enable subtle effects on germ cells and fertilisation to be evaluated.

It was concluded that there were no alternative tests available, nor were there likely to be in the medium-term future, that would enable the screening of chemicals for female reproductive toxicity with the predictivity required for labelling purposes. There was no indication that any of the individual techniques discussed at the workshop would yield a practical and generally applicable test system. In the long-term, a complex battery of in vitro tests may be devised.

Additions to in vivo testing

The draft OECD test guidelines 421 and 422 are based on a one-generation procedure, and they will predominantly provide information on fertility impairment. However, additional observations could be made, with relatively little additional effort, to provide more data on reproductive functions in the female (and in the male; see below), thereby maximising the information obtained and minimising the need for further testing. Vaginal cytology can be examined to monitor the oestrous cycle, and this assessment could be integrated into a repeated dose toxicity study on sexually mature animals (10). High quality histopathology of the ovaries should enable the detection of any adverse effects on specific populations of ovarian cells, and of the arrest of folliculogenesis, for example. Again, this could be carried out as part of a repeated dose study.

Male fertility

There is particular interest in the development of alternative approaches for assessing male reproductive toxicity because laboratory animals may not be good models. The human male has a relatively low sperm count; the number of sperm per ejaculate is typically only between two-fold and four-fold higher than that at which fertility is significantly impaired. In contrast, the number of sperm in a rat or rabbit ejaculate is many times (up to 1440-fold) that which will produce maximum fertility. It has been reported that epididymal sperm count can be reduced by as much as 90% in the rat without significantly affecting fertility. Consequently, the effects of chemicals on the fertilities of animal models may be insensitive indicators of human reproductive hazards. Studies on male (cf. female) reproductive toxicity are also aided by the ready availability of human target cells.

Alternative approaches

The production of spermatozoa from stem cells is a complex process taking about five weeks in mice and 11 weeks in humans. It is generally believed that the Sertoli cell, which is intimately associated with the developing germ cells, orchestrates spermatogenesis. Chemicals can disturb normal spermatogenesis by direct interaction with targets within the testis itself, or by indirect mechanisms, such as interference with hormonal stimulation or alterations in blood supply.

Overall, it was concluded that the current situation relating to alternative tests for male reproductive toxicity is much the same as for the female; it is not possible a priori to mimic the whole of the reproductive cycle in vitro, but several components can be studied individually. Although these components have been more extensively utilised for toxicological studies than is the case for the female, they still do not represent a viable alternative. Unlike in the female, germ cells are produced from stem cells throughout mature life in the male. Thus, it should be easier to devise culture systems which are able to support the whole of spermatogenesis. Although this is not yet possible, there is a considerable amount of research in this area, and the production of a suitable system for undertaking reproductive toxicology studies would be highly beneficial.

Several testicular cell types can be maintained in culture, either alone or in combination; these include Sertoli-germ cell co-cultures, Sertoli cell-enriched cultures, germ cell-enriched cultures, Leydig cell cultures, and Leydig-Sertoli cell co-cultures. All of these systems have been used successfully to study specific features of testicular toxicity (7). Primary cultures of testicular cells retain many of the differentiated characteristics of their in vivo counterparts, but they are inherently variable and generally have only a limited lifespan, which necessitates the frequent preparation of fresh cultures. The ability to study individual cell populations from a heterogeneous target organ such as the testis is a powerful tool for probing mechanisms of toxicity. However, the
loss of interactions with other cell types that influence their function in vivo is a serious limitation to their use for screening purposes.

**Additions to in vivo testing**

Additional measurements could profitably be incorporated into the current in vivo testing protocols. High quality histopathology of the testes and epididymes would enable effects on specific cell populations to be evaluated. Alternatively, recently developed flow cytometric techniques could be utilised. Flow cytometry offers particular advantages for cell analysis, including: a) rapid cell measurements at rates of up to 10,000 cells per second; b) simultaneous, correlated measurements of up to five parameters per cell; c) unbiased selection of the cells measured; d) high statistical significance due to efficient measurements on large numbers of cells, providing confidence in the detection of rare effects; and e) the ability to sort physically, from a population of cells, large numbers of single cells that can be further evaluated for other biochemical or morphological parameters. One of the flow cytometric methods developed for detecting alterations in spermatogenesis utilises an acridine orange staining protocol (11), which enables the simultaneous measurement of cellular DNA content, and RNA content or stainability. This procedure provides a rapid assessment of up to eight different testicular cell populations.

**Semen analyses — human and animal**

Semen analysis can provide an acceptable alternative to histopathology since there is considerable overlap in the effects which the two procedures are able to detect (12). Histopathological examination is slightly more advantageous for detecting effects on spermatogenesis, whereas semen analysis or a mating trial would be better for detecting effects on post-testicular stages, such as sperm maturation and function. There are several techniques available for monitoring sperm motility, motion and morphology, and various other aspects of semen composition; some of these analyses can be automated, for example by using videography. It was felt that these approaches should be supported. In particular, the development of new endpoints, and the comparison of all of these parameters with the fertilising ability of sperm, were considered to be important.

Such approaches could be used for both human and animal semen analyses. In addition, it was felt that the direct addition of test chemicals to semen samples in vitro may be a valuable approach, given the availability of human material; research into the types of effects that can be detected in this way should be encouraged.

Sperm head morphology and sperm chromatin structure are candidate parameters for further evaluation. It has been shown that fertilising capacity declines with increasing proportions of abnormal head morphology, and there is a high correlation between chemical-induced sperm head morphological abnormalities and altered sperm chromatin structure in the mouse (13). The sperm chromatin structure assay is a flow cytometric method that detects damage to chromatin as an increased susceptibility to acid-induced or heat-induced DNA denaturation in situ (14); when complexed with native double-stranded DNA, acridine orange fluoresces green and can be distinguished from the complex with denatured DNA, which fluoresces red.

**Embryotoxicity**

Over the past 15 years, more than 30 different systems have been proposed as tests for developmental toxicity, but the vast majority of these have been used by only one or two laboratories. The systems fall into four categories: a) established cell lines; b) primary cell cultures; c) non-eutherian embryos (both non-vertebrate and vertebrate); and d) cultured mammalian embryos or primordia. The workshop participants agreed that there is little need for any more systems to be devised, since they are unlikely to be more than variants of existing methodologies. Instead, further work should be directed toward improving the existing systems.

**Validation**

The general strategy for validating developmental toxicity screens was discussed. This is a major topic, and it was decided that it was inappropriate to attempt a full critique in this report. The selection of test substances to be used for validation purposes is critical, and efforts should be made to compile a list of suitable chemicals, which meet the criteria outlined by Balls et al. (15).
Other key issues that must be considered include: a) the manifestations of developmental toxicity (retardation, functional impairment, malformation, death) to be used with respect to the in vivo reference data and in the alternative test system; b) the ability to distinguish general toxicity from specific developmental effects; c) the acceptability of positive/negative classifications in comparison with quantitative estimates; d) the relationship between potency in vivo and in vitro; e) acceptable levels of false negatives and false positives; and f) objective definition of positive and negative results with statistical reliability.

Cell lines and embryonic stem cells

Established cell lines that have been utilised for developmental toxicity screening include human embryonic palate mesenchymal (HEPM) cells (16), mouse ovarian tumour (MOT) cells (17), and neuroblastoma cells (18). Although various cellular properties can be investigated (for example, proliferation, attachment and differentiation), most methods incorporate only one or two. The results of a blind trial with a dual HEPM/MOT approach, published in 1988 (19), showed an unacceptably high level of false positives (>50%) and, in general, this was attributed to the very simple nature of the model systems. It is the opinion of the workshop participants that super-confluent “micromass” cultures, which support a wider range of cellular processes (see below), are inherently superior to isolated cell systems. Nevertheless, a clear drawback of the orthodox micromass approach is that it requires the primary culture of cells isolated from embryos. Even if the sacrifice of mammals is avoided by using avian embryos, the need to isolate fresh cells is a practical limitation. Until recently, there had been little success in using established cell lines in micromass cultures. However, the discovery that cells of the mammalian blastocyst can be used to create totipotent embryonic stem (ES) cell lines has gone some way to addressing this problem.

Mouse ES cells can be maintained in an undifferentiated state in the presence of feeder layers and/or purified leukemia inhibitory factor (LIF). On removing the LIF, ES cells differentiate into a variety of cell types, depending upon the culture conditions. For example, in the mouse, muscle cell differentiation from ES cells reflects myogenesis in vivo (20), and the development of haematopoietic cells from ES cells parallels haematopoiesis in the developing embryo (21).

ES cells offer several new approaches with respect to screening for embryotoxicity in vitro, enabling the use of differentiating permanent embryonic cells. Cytotoxicity (22) and effects on differentiation (21, 23) have been used as endpoints in embryotoxicity tests with ES cells; inhibition of the differentiation of ES cells in a micromass culture has proved to be a particularly promising assay under routine testing conditions (23). However, determination of the two essential features of embryotoxins (that is, inhibition of differentiation combined with a higher sensitivity of embryonic cells than adult tissues to cytotoxic damage) in a single assay procedure has not yet been undertaken with ES cells.

The full potential of ES cells has yet to be realised. Automated immunological methods, for example, analyses using the fluorescence-activated cell sorter (FACS), are now being employed to study the differentiation of ES cells in the presence of developmental toxicants. Recently, Hooge & Ooms (24) have described an in vitro developmental toxicity assay, employing embryonal carcinoma cells, which uses FACS analysis for rapid quantification. Immunological staining techniques, and studies of gene expression in differentiating ES cells, may enable the identification of effects on markers which are characteristic of essential steps in development.

ES cells are routinely used in the production of transgenic mice, and methods to introduce targeted mutations and reporter constructs are well established. Transgenic markers could be devised to simplify the endpoints used in a particular toxicity test, and to enable the automation of such assays. It is possible that ES cell lines isolated from primates could also be utilised. The workshop participants supported the continued development of these approaches.

Micromass cultures

The micromass test is intended to identify substances which induce malformations (i.e. teratogens) resulting in embryotoxicity. The micromass culture technique was devised by Umansky (25), who reported that, when cells
from the undifferentiated mesenchyme of early chick embryo limbs were cultured in small volumes at high density, they formed numerous small foci of differentiating chondrocytes within a background of apparently undifferentiated cells. It has since been documented that cell adhesion, movement, communication, division and differentiation all occur in micromass cultures (26). In principle, the micromass test is based on detecting the ability of a particular chemical to inhibit the formation of foci. Embryonic limb or central nervous system (CNS) cells (usually mid-brain, which form foci of neurons) from chick, mouse or rat can be used (27–29). The technique has subsequently been modified for use with 96-well microtitre plates (28, 30). Cells can be exposed either directly in culture or transplacentally prior to culture (31).

Several structure-activity studies have shown that the micromass test can distinguish between teratogens and non-teratogens within a particular chemical class, for example retinoids (32) and triazole antifungals (33). In some cases, organ-specific, species-specific and strain-specific toxicities have been modelled in micromass cultures. For example, ethylendieniourerea is more toxic to mid-brain than to limb cultures (34), and is more toxic to rat CNS than to mouse CNS (35), while the effects of retinoic acid differ in limb micromass cultures prepared from Sprague-Dawley and Wistar rats (36).

Differentiating cells in both mid-brain and limb cultures express cytochrome P450 isoenzymes (37, 38) and, at least in the case of limb cells, these are able to metabolise chemicals such as phenytoin and cyclophosphamide to their toxic metabolites.

Validation studies using chemicals from a variety of classes indicate that the percentage of teratogens detected with micromass cultures may vary between 60 and 90%, and that the percentage of non-teratogens identified correctly may vary between 89 and 100% (26, 29, 39, 40). It is possible that much of this variation is accounted for by differences in the exact methodology used, and none of these studies were considered to be definitive by the workshop participants.

Currently, the results of an interlaboratory blind trial are being prepared for publication (Oliver Flint, personal communication). An initial study with 24 compounds indicated little interlaboratory variability, despite some differences in the test protocols employed. Test substances were classified into four groups based on their effects in vivo: Group I — teratogenic in all species tested at a dose that does not induce maternal toxicity; Group II — teratogenic in some (but not all) species at a dose that does not induce maternal toxicity; Group III — teratogenic in some species, but only at doses which induce maternal toxicity; and Group IV — not teratogenic in animal tests. In the past, various definitions were used for a positive result in micromass cultures. The approach being used in the blind trial is that results indicate teratogenic potential if any two of the following conditions are met: a) the IC50 values (that is, the concentrations inhibiting cell survival or differentiation by 50% relative to the controls) in limb or mid-brain cultures are less than 100μg/ml; b) the IC50 value for the inhibition of cell survival is at least twice that for the inhibition of cell differentiation in limb or mid-brain cultures; and c) the IC50 value for the inhibition of cell differentiation in one culture type is at least twice that for the inhibition of cell differentiation in the other culture type.

Based on these criteria, the four Group I chemicals were all positive and seven out of eight Group IV chemicals were negative. False negatives occurred among those substances in Groups II and III but, with the exception of diphenylhydantoin, these compounds were not considered to represent a risk with respect to human teratogenicity under normal conditions of exposure. On the basis of these results, it was concluded that the micromass test was able to identify teratogens having high potencies, did not over-predict the teratogenic potentials of compounds with minimal risks of inducing teratogenic effects in humans, and had a low rate of false positives.

Thus, micromass cultures represent robust test systems for studying potential teratogens. There is probably more experience with this system than with any other alternative method for reproductive toxicology studies; its limitations and the correct interpretation of the data obtained are now well understood. It is recommended that the micromass method be included in a comparative trial, to determine its applicability relative to several other in vitro systems which are available.
Embryos of anurans and lower order species

Numerous tests which use embryos of submammalian vertebrate and invertebrate species for detecting the teratogenic potentials of chemicals have been described. The organisms which have been used include fish, frogs, crickets, Drosophila, brine shrimps and slime moulds. Several of these are currently being used extensively as models for investigating mechanisms of development. Although it is clear that there is substantial evolutionary conservation of basic mechanisms (for example, in the transcription factors used to encode regional specification), there are, nevertheless, significant differences between species. Since any stage or component of development is a potential target for toxicants, the existence of species differences is a strong argument in favour of using vertebrate models for predictive screening. However, it is acknowledged that sub-vertebrate systems may have applications in ecotoxicological monitoring.

Of the non-avian vertebrate systems available, the workshop participants concluded that only the system using the South African clawed frog, Xenopus laevis, was worthy of further attention. The other systems were rejected on the basis of insufficient practical experience and validation. The Xenopus system, or FETAX (frog embryo teratogenesis assay: Xenopus), has undergone limited validation using about 40 different substances (41-43). Its overall accuracy in predicting teratogenic potential has been claimed to be 79-83%. FETAX is low-cost and rapid, and uses a species commonly maintained under laboratory conditions. However, there are some limitations of the assay, which relate to the aqueous solubility of test substances and their penetration into the embryo, the relative lack of data pertaining to its validation, and the small number of laboratories that have used the system. Nevertheless, it is recommended that FETAX should be included in a comparative trial of alternative tests for developmental toxicity.

Avian embryos

Avian embryos are, in many respects, ideal alternatives for developmental toxicity screening: they are non-eutherian, and non-mammalian but warm-blooded; there is an extensive literature on their embryology; and they are currently being used widely as models in developmental biology studies. Nevertheless, avian embryos are seldom used in testing for embryotoxicity/teratogenicity. This is probably due to the recommendation of a World Health Organisation report published in 1967 (44), which characterised the chick embryo as being too sensitive and variable. At that time, chick embryo studies used blind yolk-sac injections which resulted in poorly reproducible results. However, several new approaches have now been developed, and the workshop participants concluded that it is time to re-evaluate the chick embryo as a possible screen.

The chick embryotoxicity screening test (CHEST) was devised by Jelinek and co-workers (45-47), and has been used extensively in their laboratory, but not elsewhere. CHEST is divided into two phases: CHEST I, which is performed on incubation day 1.5; and CHEST II, which is performed on days 2, 3 and 4. It uses eggs opened by the windowing technique, to allow accurate selection of embryos and the targeted administration of very small amounts of test substances (both of which are prerequisites for reproducible results). Intra-amniotic injection eliminates the problem of continuous exposure of the embryo, because the test substance is readily distributed to the extraembryonic compartments. Growth retardation, malformation and death can be monitored. Dose-response and stage-response relationships, and malformation spectra, are easily determined. The results obtained from testing over 130 compounds have been published (48-51).

It is recommended that CHEST should be included in a comparative trial. In the meantime, ways should be devised to analyse CHEST results so that they are comparable with those derived from other test systems.

Another test system has been devised that uses chick embryos in culture, rather than in ovo (52). The explanted chick embryo system shares some of the advantages and disadvantages of mammalian whole-embryo cultures (see below). However, its proponents suggest that the culture methods are easier and less expensive, and that it does not require mammals to be sacrificed. Recently, it has been shown that chick and mammalian systems gave comparable results for 12 test chemicals (52). The workshop participants concluded that a major benefit of the chick embryo was that the whole of embryogenesis could be
studied in ovo, and that this outweighed the advantages of using a culture system.

**Mammalian embryo culture**

Mammalian embryos can be maintained in culture for short periods throughout the phase from fertilisation to the end of organogenesis. For the purposes of toxicity testing, the period from the end of gastrulation to mid-organogenesis has been investigated extensively. Screening systems using mouse (for example, 53) and rat (for example, 54) embryos have been proposed, and the culture of rabbit embryos has recently been optimised (55).

In brief, head fold or early somite stage embryos are dissected free from maternal tissue, parietal yolk sac and Reichert's membrane, leaving the visceral yolk sac and ectoplacental cone intact. The conceptus is cultured in medium under defined gassing conditions for 24-48 hours, usually in a roller bottle system. A variety of media have been used, all of which contain a high proportion of serum. Rat serum is most common (for example, 56), but mouse (57), rabbit (55), cow (58), monkey (59) and human sera (for example, 60) have been used. The test compound can be added to the cultures for defined periods or for the entire culture period. Metabolic activation systems can be incorporated, including the addition of S9 or microsomal fractions of liver from different species, co-culture with hepatocytes, sequential hepatocyte/whole-embryo culture, and the addition of serum from treated animals or humans (61–66).

At the end of the culture period, a number of endpoints can be measured, including: a) effects on the development of the visceral yolk sac vascularisation and circulation; b) effects on haematopoiesis, embryonic growth (for example, size, and protein and DNA contents) and differentiation (number of somites, morphological score); and c) dysmorphogenic effects (53, 54, 67–69). The interpretation of the results obtained takes into account adverse effects on yolk sac development, embryonic growth and differentiation, as well as specifically on dysmorphogenesis.

Validation studies have been published by Schmid and colleagues (70–72), and an interlaboratory validation study is ongoing (73–75). Results of a validation study on different culture systems have been published by Kucera et al. (52), in which six pairs of coded compounds were tested in chick and rat embryo cultures and in brain cell aggregate cultures. Bechter et al. (76) reported an excellent agreement between in vivo and in vitro data for a series of retinoids.

The workshop participants concluded that mammalian whole-embryo culture systems are well developed in vitro tests for the detection of potentially teratogenic compounds and for the elucidation of mechanisms of teratogenicity. Such in vitro systems have been used in many laboratories, both academic and industrial, and have proved to be valuable tools. Whole-embryo culture enables the detection of dysmorphogenesis in many organ primordia, and the comparison of specific dysmorphic effects with general adverse effects on growth and differentiation. In addition, it enables the potencies of structurally related compounds to be ranked. Test compound and/or metabolite concentrations in the culture medium and in conceptus tissue can easily be monitored.

However, the system has clear limitations: it is relatively complex, covers only a part of organogenesis, and requires a high level of technical skills. The test can be costly, and it uses mammalian tissue and serum. Whether this is justified with respect to its use as a screen should be evaluated by including it in a comparative trial with other simpler in vitro systems. It is likely that these different systems will provide complimentary information, and selection of the appropriate test for a particular application should be made on a case-by-case basis.

**Toxicokinetics and Metabolism**

The production of a direct effect on the developing organism depends on the concentration–time relationship of the chemical and/or its active metabolite(s) in the target cells. Therefore, toxicokinetic and metabolism studies are of crucial importance for the design and interpretation of both in vitro and in vivo developmental toxicity studies (77, 78). In vivo, the target concentrations are dependent on maternal absorption of the compound, its distribution, metabolism and excretion, and, in particular, its placental transfer and distribution within the conceptus. Toxicokinetic studies are just as important in vitro, when the presence of the compound and its stability in the culture
medium must be verified, along with an assessment of its transport to, and uptake by, the tissues/cells in culture, its metabolic activation, and its cellular distribution.

There are many parallels between the toxicokinetic principles of in vivo and in vitro studies. For example, the concentrations of a compound and its metabolites in plasma in vivo relate to those in the culture medium in vitro, and the crucial importance of the levels reached in the target tissue holds for both systems. Protein binding in maternal plasma is mirrored by protein binding in the culture medium; both are very important with regard to the placental transfer of some classes of compounds (for example, acidic compounds; 79, 80).

Toxicokinetic studies are the key to interpreting results obtained in vitro, and for extrapolating these to the in vivo situation. Therefore, it is strongly recommended that the concentrations of chemicals/metabolites in culture media are measured, in order to verify the concentration of compound added (this is of particular importance with very lipophilic compounds), and to check the stability of the added compound and assess the production of metabolites during the culture period. The activities of added metabolising systems, such as liver homogenate fractions, isolated enzymes and hepatocytes, can be assessed by analytical techniques. Measurement of the compound in the cultured tissues/cells is critical, so that the target concentration needed to yield an effect can be determined. Such measurements are especially important if little or no activity is observed in vitro, so that false negatives can be excluded. In such cases, the level of exposure of the target tissue in culture must be known, to decide whether the lack of effect results from a lack of exposure or from the low intrinsic activity of the chemical tested.

Toxicokinetic parameters often differ drastically between the in vivo and in vitro situations. For example, in vivo drug levels can fluctuate markedly between doses because of the short half-lives of many chemicals, and high concentration peaks alternate with low or negligible drug levels. In contrast, in in vitro studies, the chemical is added to the culture medium, and persists there for extended periods of time unless it is degraded by hydrolysis or by enzymes present in the culture medium. To mimic the in vivo situation in culture, "pulse" experiments are recommended, in which the drug is added for a limited time only and then the medium is replaced. Higher chemical concentrations are probably needed with these short-exposure experiments, but the effects produced may be more specific since the non-specific effects (for example, general retardation) may be suppressed.

Toxicokinetic studies are of great importance for the interpretation of structure-activity relationships, as has been demonstrated with retinoids (81, 82) and with valproic acid analogues (77, 79, 80). The activities of substances which are closely related structurally can be interpreted only if there is information about the exposure of the target tissue. Rapidly metabolised substances, and compounds which are poorly distributed to/taken up by the target cells, may never reach effective concentrations in the target tissue, and thus may be inactive in certain systems (80). Due to differences in placental structure and function between species, such results may not necessarily be applicable to the human situation. It is recommended that pairs of enantiomers which have been demonstrated to be similar with regard to their toxicokinetic behaviour, including their distribution to the target tissue, but which exhibit different activities (80), are used for evaluating and validating in vitro systems. Such enantiomeric pairs provide very specific probes for studying whether a particular in vitro method truly reflects the situation observed in vivo (83–85).

Recommendations

The following recommendations are of equal priority and should be implemented concurrently.

1. Four in vitro systems of differing complexities and design, Xenopus embryos (FETAX), chick embryos (modified CHEST), micromass cultures and mammalian whole-embryo cultures, are well-established methods which are capable of detecting many substances likely to exert potent effects on the physical development of the embryo. It is recommended that these four systems be compared, and that ECVAM is responsible for coordinating this validation study and for acting as
a repository for test substances, data and protocols.

2. Methods using mammalian embryonic stem cells should be further developed, optimised and validated. Such systems have the potential to replace cellular methods based on primary cultures, and can incorporate transgenic markers to enhance their applicability and practicality.

3. Toxicokinetic data should be obtained at the outset of all toxicity testing, both in animal and in alternative tests. Such data can significantly enhance experimental design and aid the subsequent interpretation and extrapolation of all test results.

4. OECD draft test guidelines 421 and 422 should be adopted for the reproductive and developmental toxicity testing of base-set chemicals, with manufacturers being given the option of using either guideline. In both cases, histopathology of the testes and epididymes, and of the ovaries, must be of a sufficiently high quality to detect changes related to effects on reproduction. Vaginal cytology, sperm analysis and cytometric analysis of spermatogenic cell types are recommended as useful additions. All of these direct methods can be used to supplement other toxicity tests, such as 28-day repeat dose studies (OECD test guideline 407), to reduce animal use.

5. It is recommended that base-set chemicals should be labelled on the basis of these enhanced OECD test guidelines. The extent to which any further testing may be required depends upon many factors and can only be decided on a case-by-case basis.

6. The OECD test guidelines are, in essence, the core of comprehensive reproductive toxicity tests; the animal procedures outlined have been used very widely and so they need no further validation. To demonstrate their value conclusively, we recommend that all available results be collected and evaluated, and that this be continued when/if the proposed regulations are implemented. Thorough analysis of existing test data and those produced in the future would enable the informative components of current methods to be identified, thus preventing animals being used to produce uninformative results. We strongly believe that manufacturers should be encouraged to participate in the parallel development and validation of animal procedures and alternative tests.

7. For reproductive toxicity testing, the current alternative methods can only be used to evaluate a few components of integrated reproductive function, particularly in the female. Methods using human semen for testing the effects of chemicals on mature sperm should be refined and validated.

8. It is proposed that another EU-sponsored workshop be convened in four years time, to evaluate progress and reassess the potential of alternative tests. We believe that the objectives of our recommendations can realistically be met within this period.

References


cokinetie and structure-specificity. Archives of Toxicology, Suppl. 16, 118–127.

