

WellBeing International

## WBI Studies Repository

10-2007

# The Development of New Concepts for Assessing Reproductive Toxicity Applicable to Large Scale Toxicological Programmes

S. Bremer

*European Centre for the Validation of Alternative Methods*

C. Pellizzer

*European Centre for the Validation of Alternative Methods*

S. Hoffmann

*European Centre for the Validation of Alternative Methods*

T. Seidle

*European Centre for the Validation of Alternative Methods*

T. Hartung

*European Centre for the Validation of Alternative Methods*

Follow this and additional works at: [https://www.wellbeingintlstudiesrepository.org/acwp\\_arte](https://www.wellbeingintlstudiesrepository.org/acwp_arte)



Part of the [Bioethics and Medical Ethics Commons](#), [Laboratory and Basic Science Research Commons](#), and the [Research Methods in Life Sciences Commons](#)

---

### Recommended Citation

Bremer, S., Pellizzer, C., Hoffmann, S., Seidle, T., & Hartung, T. (2007). The development of new concepts for assessing reproductive toxicity applicable to large scale toxicological programmes. *Current pharmaceutical design*, 13(29), 3047-3058.

This material is brought to you for free and open access by WellBeing International. It has been accepted for inclusion by an authorized administrator of the WBI Studies Repository. For more information, please contact [wbisr-info@wellbeingintl.org](mailto:wbisr-info@wellbeingintl.org).



# The Development of New Concepts for Assessing Reproductive Toxicity Applicable to Large Scale Toxicological Programmes

S. Bremer, C. Pellizzer, S. Hoffmann, T. Seidle and T. Hartung

*European Centre for the Validation of Alternative Methods, Institute for Health and Consumer Protection, Joint Research Centre, European Commission, Via Fermi1, 21020 Ispra (VA), Italy*

## CITATION

Bremer, S., Pellizzer, C., Hoffmann, S., Seidle, T., & Hartung, T. (2007). The development of new concepts for assessing reproductive toxicity applicable to large scale toxicological programmes. *Current pharmaceutical design*, 13(29), 3047-3058.

## KEYWORDS

REACH, reproductive toxicity, prevalence, alternative methods, validation, ReProTect

## ABSTRACT

Large scale toxicological testing programmes which are currently ongoing such as the new European chemical legislation REACH require the development of new integrated testing strategies rather than applying traditional testing schemes to thousands of chemicals. The current practice of requiring *in vivo* testing for every possible adverse effect endanger the success of these programmes due (i) to limited testing facilities and sufficient capacity of scientific/technical knowledge for reproductive toxicity; (ii) an unacceptable number of laboratory animals involved (iii) an intolerable number of chemicals classified as false positive.

A key aspect of the implementation of new testing strategies is the determination of prevalence of reproductive toxicity in the universe of industrial chemicals. Prevalences are relevant in order to be aware on the expected rate of false classification during the toxicological testing and to implement appropriate measures for their avoidance. Furthermore, a detailed understanding on the subendpoints affected by reproductive toxicants and the underlying mechanisms will lead to more science based testing strategies integrating alternative methods without compromising the protection of consumers.

## INTRODUCTION

The relationship between chemical exposure and reproductive/developmental toxicity is highly complex, involving interactions between multiple organs and organ systems in both parents and offspring at a variety of different time points and life stages. The inherent complexity of the vertebrate reproductive system represents a significant challenge to the development of *in vitro*, *in silico* and “omic,” as well as other non-animal technologies aimed at reducing, and ultimately replacing, animal use in this area of toxicology [1]. This natural incorporation of emerging scientific opportunities has to be seen also in the context of the increasing need to review traditional approaches in toxicology, as attempted in the emerging evidence-based toxicology [2].

However, not only animal welfare considerations prompt the need to substitute traditional approaches by new testing strategies: Pronounced inter-species variances have been described showing not more than 60% correlation between different laboratory mammalian species in the area of developmental toxicity [3, 40]. There is no reason to assume that any species does predict humans better than e.g. mice predict rat developmental toxicity of a given chemical. The “precautionary” response of regulatory toxicology was to test in more than one laboratory animal species, in order to reduce the 40% missed potentially developmental toxic substances. However, inevitably this increases on the other side the already 40% false-positive classifications. Whether we can afford this substantial overlabelling especially in high-production volume chemical evaluation programmes shall be discussed.

There is also in various respects an economical dimension: reproductive toxicity testing was already identified as the driving factor of costs (and animal use) and also as the driving factor for the new European chemicals legislation REACH [4, 5]. This is due to the fact that tests like the two-generation study are among the most costly tests and at the same time requiring up to 3.200 animals (two generation study) per substance. In the light of the finalization of the legislation, a reassessment of these figures for reproductive toxicity appeared to be appropriate.

Often overlooked, reproductive toxicity testing has not been developed for, nor been largely applied to, chemicals in general. It has predominantly been used for pharmaceuticals and pesticides. Although introduced for chemicals several decades ago, this held true only for new chemicals at a certain production volume. However, very few new chemicals are produced in volumes triggering such testing. Thus, experience of the predictive value and performance in general for ordinary chemicals is more than limited. So are also the laboratory capacities available to carry out testing.

Testing requirements for reproductive toxicity in various regulatory sectors (e.g. pharmaceuticals, agrochemicals, industrial chemicals) should reflect current knowledge concerning the relative contribution of different substances to the overall prevalence of adverse reproductive outcomes. In this context, only about 2% of birth defects have been attributed to environmental chemicals as a causal factor [6].

A detailed understanding of target tissues and biological mechanisms of reproductive toxicants is necessary to effectively implement the use of alternative methods for highly complex endpoints for which a one-to-one replacement cannot be achieved. Despite the complexity, many regulations already request that the principles of the (3Rs) are respected [7] in order to ensure that the most recent toxicological methods are used to reduce the number of animals while not compromising the safety of consumers and patients (e.g. as in Europe’s new chemical legislation “REACH” requires in its Article 1 “...*This Regulation should also promote the development of alternative methods for the assessment of hazards of substances...*” or the draft of revision of Annex II and III to Council Directive 91/414/EEC for plant protection products is asking for “...*study designs that have taken reduction, refinement and replacement of animal tests into consideration*”). The seventh amendment to *Directive 76/768/EEC* on cosmetics from 2003 foresees a complete marketing ban on cosmetics ingredients tested for reproductive toxicity on animals in 2013. These legislations require study designs that take reduction, refinement and replacement of animal tests into account when performing hazard assessment of substances. However, there is still a strong reluctance by regulatory authorities with regard to the use of validated *in vitro* tests in a context of testing strategies due to their limitations with regard to coverage of sub-endpoints and xenobiotic metabolism, but also due to a limited confidence in the reliability of formal validation studies. Integrated, “intelligent” testing strategies are necessary to make optimal use of the so called “partial replacement” methods. In the following, we would like to provide a short overview of the current status of reproductive toxicity testing, its drawbacks when applied to large scale toxicological programmes and to

introduce new concepts that make use of a more scientific hazard assessment using new technologies and existing knowledge.

## **CLASSICAL APPROACH TO REPRODUCTIVE TOXICITY ASSESSMENT**

Currently the treatment of one or more generations of rodents with a test chemical is the most common approach for identifying chemically induced adverse effects on reproduction. For evaluating developmental toxicity, test guidelines were designed to detect malformations in the developing offspring together with parameters such as growth alterations and prenatal mortality.

A variant on these “definitive” studies are the shorter and less complex “screening” tests, which combine reproductive, developmental, and (optionally) repeated dose toxicity endpoints into a single study design. From the animal welfare point of view a combination of a repeated dose toxicity study with the reproduction/developmental toxicity screening test is preferable but this requires compromises due to the exceptional physiological status of pregnant animals which are eventually not providing unbiased toxicological information on the subacute (chronic) effects.

These reduced protocols have mainly been employed in national and international programmes to gather screening-level data for chemicals [8, 9]. However, this study design has limited sensitivity and produces a high level of equivocal results which often have to be further evaluated in more “definite studies”, such as a prenatal developmental toxicity study and/or a two generation study. Due to the fact that the screening requires 560 animals per test [10], the application of this test as screening tool in its present form should be reconsidered for large toxicological programmes. An improvement of the test design in order to increase resolving power of the test and by reducing the number of equivocal results is desirable.

The current concept of reproductive toxicology testing for regulatory purposes requires extensive hazard tests analysing every possible adverse effect followed by the elimination of information not relevant to regulatory decisions such as classification/labelling and risk assessment. Only in the rare event of sufficient existing information is no further testing requested. This classical approach has been applied in the past although only relatively few new chemicals have been evaluated for their hazards.

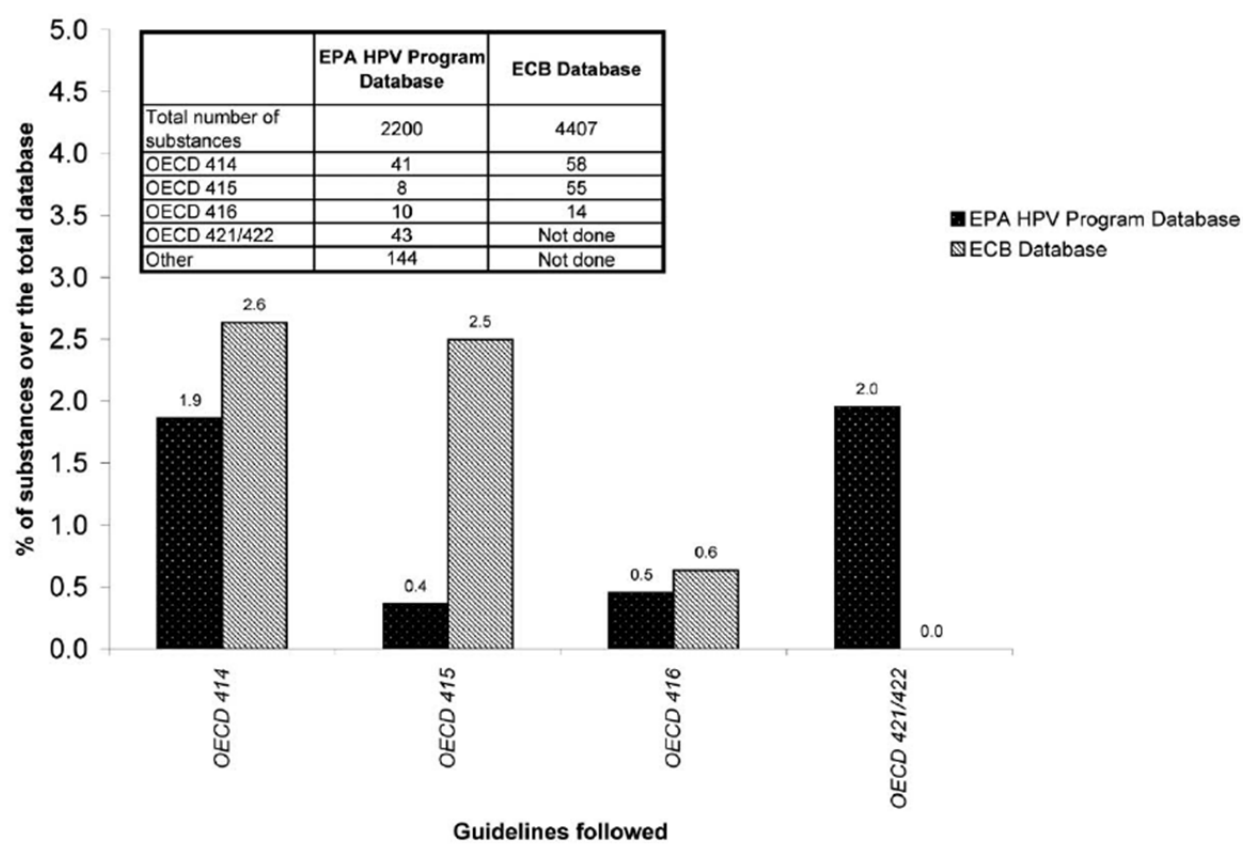
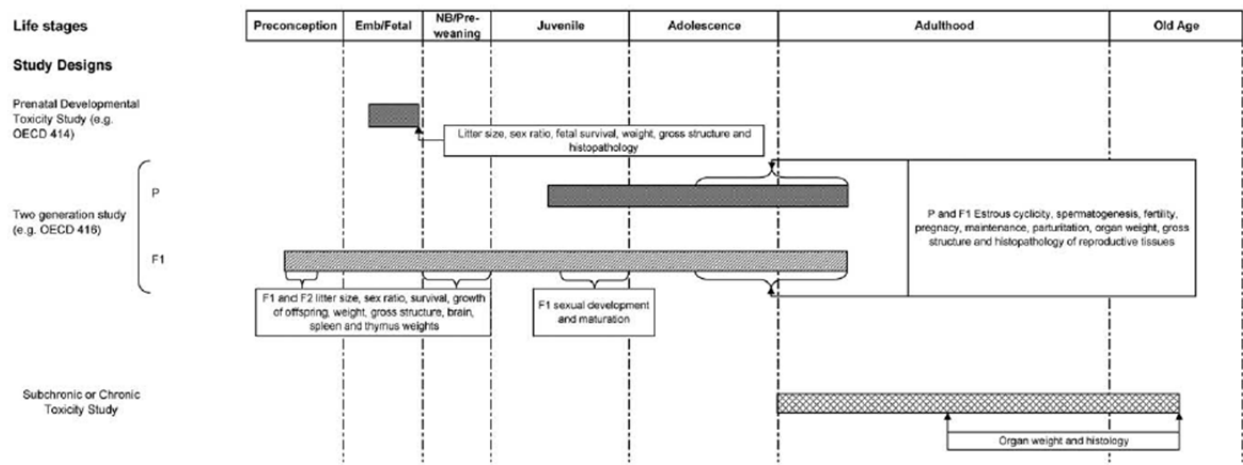
A summary of tests currently used for regulatory decision making is given in (Table 1). However, large toxicological programmes such as REACH are now requiring a reorientation of the currently used testing paradigm. The success of REACH will be endangered if regulatory bodies continue to follow a conservative approach requiring *in vivo* testing for every possible adverse effect as it is proposed in the current draft technical guidance for industry of the REACH implementation programme at higher tonnage levels [26]. The traditional concept to cover all possible endpoints is shown in Fig. (1). This approach will lead to testing between 1.665 and 5.500 chemicals in e.g. two-generation studies [10], which is to achieve due to the limited testing facilities and a lack of sufficient scientific/technical know-how. A survey carried out including 28 major independent and corporate laboratories in Europe indicated that only 11 offer two-generation studies with a capacity of 28 substances per year [27]. This total suggested a capacity to carry out about 50 parallel two-generation studies in Europe, each lasting about two years. Thus, every year 25 new substances can be included. The majority of this testing capacity is employed for drugs and pesticides. Only about three general chemicals per year have been tested in two-generation studies since the introduction of the Dangerous Substance Directive in 1981 [27]. Thus, testing of hundreds or even thousands of chemicals in the context of REACH will overwhelm available test capacities. This calls for adequate priority setting to make best use of these limited resources as well as for the use of any other mean to satisfy the information requirement by means of an integrated testing strategy.

In addition, the ethical dilemma of this conservative concept becomes clear by analyzing the impact on animal numbers for large toxicological programmes to fulfill the information requirements of REACH. Several impact assessments have addressed this problem; the most prominent ones are published by van der Jagt [5], Hoefler [10] and their colleagues. These figures have been recently reassessed [28] due to further knowledge derived from the REACH implementation projects. Unfortunately, currently only less than 5% of the toxicological dossiers of the US HPV program or in the new chemical data base of the European Chemicals Bureau (ECB) contain data for reproductive toxicity equivalent to the information level required for REACH (Fig. (2)). The most relevant tests accounting for REACH are TG 421 (560 animals/test), TG 414 (150 animals and up to 1200 fetuses/species/test) and TG 416 (3200 animals/test). There are assumptions about the number of substances falling in each tonnage range, i.e. 5.000 for 1-10 tons, 2.500 for 100-1000 tons and 2.700 above 1000 tons. If these all would require testing according to REACH, 5.7 (TG 421), 3 (TG 414) and 20.5 (TG 416) million animals would be needed. Assumptions as to available data and waiving opportunities, which cannot be detailed here, bring these 29.2 million animals down to 19.4 million. The possible impact of non-testing methods and alternative methods will depend on technical progress as well as the willingness to incorporate them in integrated testing strategies. Noteworthy, the consideration of a second species for the two-generation study, endocrine disrupter testing, as well as developmental neurotoxicity testing as a possible additional testing requirements have not even been taken into account in our calculation.

**Table 1. Overview on the Currently Used and Suggested Test Protocols**

<b>Study Type</b>	<b>Test Guideline(s)</b>	<b>Ref.</b>
1-generation reproduction toxicity study	EU Method B.34; OECD 415	[11,12]
2-generation reproduction toxicity study	EU Method B.35; OECD 416; OPPTS 870.3800	[11,13,14]
Reproduction/developmental toxicity screening test	OECD 421; OPPTS 870.3550	[15,16]
Combined repeated dose toxicity study with the reproduction/ developmental toxicity screening test	OECD 422; OPPTS 870.3650	[17,18]
Segment I and III studies for detection of toxicity to reproduction of medicinal products and toxicity to male fertility	ICH S5(R2)	[19]
Prenatal developmental toxicity (tetragenicity)	EU Method B.31; OECD 414; OPPTS 870.3700	[11,20,21]
Segment II	ICH S5(R2)	[22]
Developmental Neurotoxicity	OECD 426 (draft); OPPTS 870.6300	[23,24]
Mammalian level 5 test	Proposal to the OECD	[25]

The presented guidelines or drafts are used to identify reproductive/developmental hazards. All protocols are based on animal experiments.



**Fig. (2).** Summary of existing data from the US-EPA HPV database and the ECB database fulfilling the standard information requirements of REACH.

The expense and animal use associated with reproductive toxicity testing appear even less reasonable when one considers that reproductive toxicity is most probably an event with a low frequency in the universe of industrial chemicals. An independent expert panel of industrial reproductive toxicologists has concluded that in all likelihood, less than 5% of industrial chemicals possess properties that could be

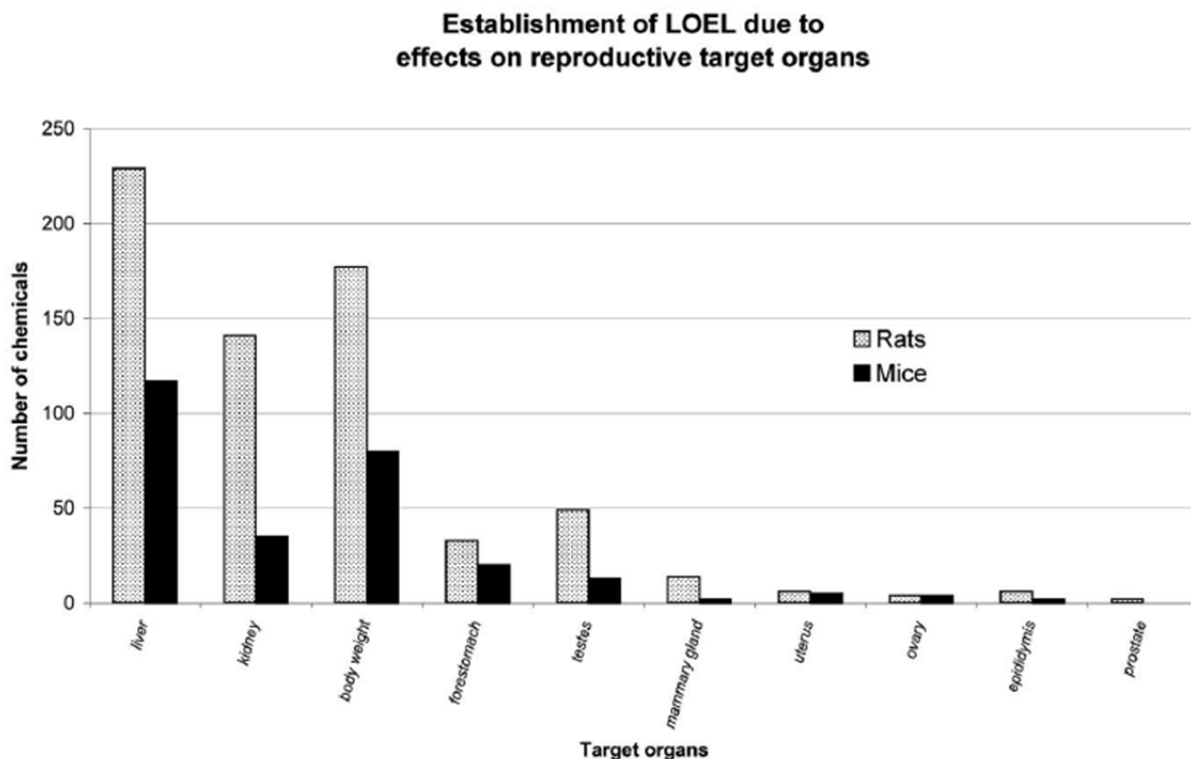
harmful to the developing child. This assumption estimate is supported by a review of the ECB New Chemicals Database, which documents that over the past 25 years, 84 developmental toxicity studies of 58 new chemicals have been performed according to OECD TG 414, leading to only 3 regulatory classifications of developmental toxicants.

This is reflected in findings of various other studies that have analyzed the prevalence of birth defects. In 3-6% of newborns, birth defects have been occurred due to various reasons such as chromosomal abnormalities, monogenetic disorders, maternal diseases, multi-factorial and spontaneous developmental disorders. Only 2% of birth defects can be associated with chemical and physical stress. This includes mainly the abuse of alcohol and other drugs [29-32]. With support of the EC DG Health and Consumer Protection (DG SANCO) Public Health Programme, the European network of population-based registries for the epidemiologic surveillance of congenital anomalies Interpretation of Prevalence Rates (Eurocat) is maintained in order e.g. to facilitate the early warning of new teratogenic exposures [33].

For the assessment of the prevalence of effects on mammalian fertility, the available database is even more limited. A query of the database for repeated dose toxicity [34] has analyzed for how many chemicals the LOEL has been established due to effects on reproductive target organs in chronic studies. Results are presented in Fig. (3). These data indicate that except for testes histopathology, effects on reproductive target organs rarely serve to establish the LOEL. It should be noted, that these are not distinguishing the route of application and the study duration; liver, kidney, and body weight are given as examples for often affected organs; forestomach for local toxicity. It needs to be stressed that the described effects do not automatically point to impaired mammalian reproduction, but only to observed histopathological effects. The prevalence of reproductive toxicity is most probably lower than this query is demonstrating. We have found by reviewing the New Chemical Database of the ECB that 15 two-generation studies have led to only one R60 classification whereas 58 one-generation studies have led to 3 classifications.

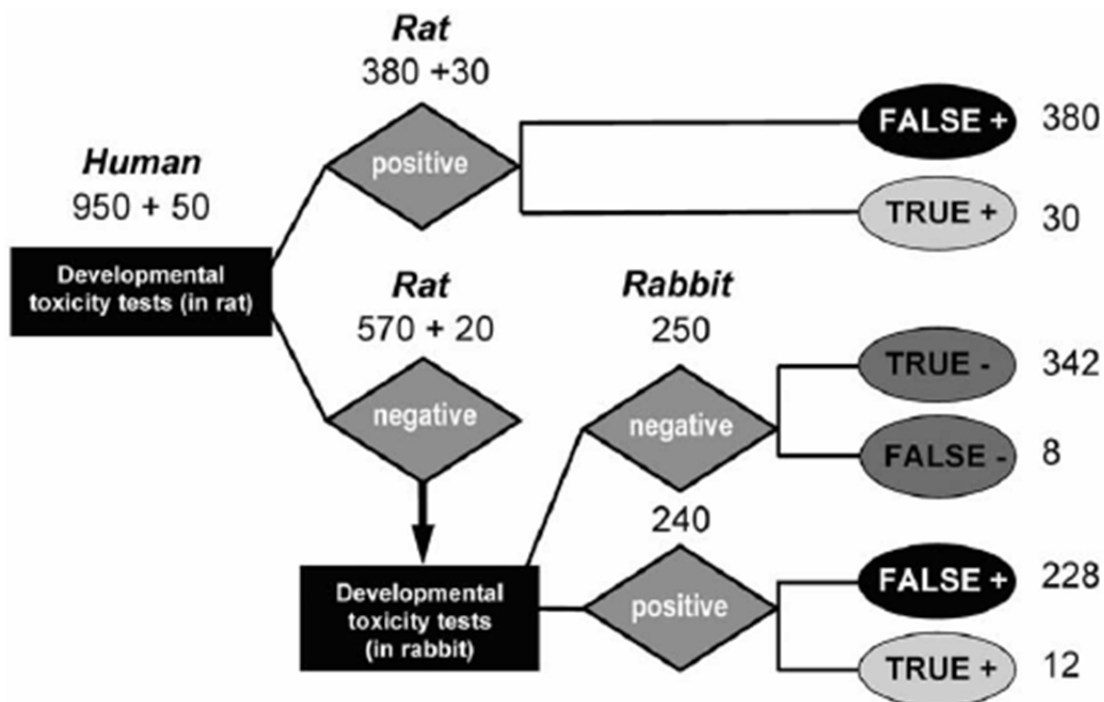
Aside from the feasibility of fulfilling the current information requirements in large scale testing programmes, the uncertainty of the testing design itself has not been assessed. Most of animal-based test methods never underwent formal validation according to internationally agreed criteria [35-37]; a firm conclusion on the sensitivity and specificity of these tests and their relevance for predicting adverse effects to humans can thus not be made. The bitter lesson of thalidomide has pointed to the need for predictive tests. The disaster has made clear that results of developmental toxicity testing in one laboratory species cannot easily be translated to humans if the predictive capacity of a test is not defined. Several studies have shown that no single animal species is more predictive to human teratogens than another, and no species is more sensitive to teratogens in general [38-40]. Hurr et al. [40] have demonstrated by analyzing 91 veterinary drugs that no single species (rat, rabbit, or mouse) was capable of detecting more than 61% of the teratogens. By using two species, the hazard of all drugs could be detected in the study of Hurr. However, this study should be interpreted with caution since Schardein [31] has provided an extensive study in which several hundreds of chemicals have been assessed for their interspecies variations. In order to overcome the drawback of low sensitivity, regulatory bodies request testing in a second species. It should be stressed that the sensitivity of the test design requesting two species is still not known. But the consequence of requesting two species is dramatic: By assuming a prevalence of maximal 5% for developmental toxicity in the universe of industrial chemicals and by requesting additional testing in another species in case of a negative first study, the number of animals for developmental toxicity testing will be nearly doubled. In addition, a side effect of requesting a second species which is often overlooked in the current testing practice but which will have a high impact on large testing programmes is the increase in the rate of falsepositives, and therefore the unwanted restrictions of valuable substances. A simple calculation might illustrate the consequences of the traditional approach.

We assume that a maximum of 5% of chemicals exhibit developmental toxicity in humans; this means that in 1000 chemicals we have to identify 50 toxic chemicals. Applying a first species (typically rat) which correlates by 60% with humans implies that we identify 30 of these 50. A second species (typically rabbit) applied to the 20 missed ones, will identify 12 correctly. This means that we still miss one in eight developmental toxicants. The major problem, however, becomes evident for the harmless substances in humans: 40% from 950 innocent substances will be labelled false-positively in rat, i.e. 380 chemicals. When testing in the second species, whose test results are assumed to be conditional independent from the results of the first species, the remaining 570 negative chemicals, another 40% (228) will be classified false-positively. Thus, in total this sums up to 608 false-positive classifications in order to identify 42, still missing 8 true ones. This classification will result either in restrictions of use, abandoning of substances, or costly follow-up studies to demonstrate that these findings do not translate to humans. The calculation is illustrated in Fig. (4). whether we can afford such sacrifice is a question beyond the remit of this article. Presenting this calculation on occasion of a number of scientific meetings resulted in very different reactions: While there was little concern about the underlying assumptions (maximum 5% real-positives and 60% interspecies correlation), practitioner's reactions varied from agreement with regard to experienced classifications to a complete disbelief. Most probably, the correlation between species is indeed better, since the number used is based mainly on substances which are positive in one species and have been retested in other species, while the majority of substances should be negative in all species. However, the prevalence of developmental toxicants is most probably even lower (2-3%, as indicated from the analysis of various databases). Assuming 2,5% prevalence, even 70% inter-species correlation still results in 497 false-positives and 80% correlation still in 351 false-positives. Thus, even a drastically increased assumption with regard to inter-species correlation leaves us with an enormous fraction of false-positives in order to identify about 24 real positives. The positive predictive value, i.e. how sure can you be of a positive test result, is thus somewhere between 4 to 8%.



**Fig. (3).** Contribution of different (reproductive) organs to establishing LOAEL in repeated dose toxicity tests in mice and rats (n=329 chemicals/rats; 203 chemicals/mice).





**Fig. (4).** The presented scenario is based on two assumptions: (i) the prevalence of developmental toxicity in the universe of industrial chemicals is 5% and the correlation between species is 60% [38]. However, the current testing scheme will result under this assumption to a false-positive classification of more than 600 chemicals since clear positive results in the first species will lead to regulatory decisions.

Other interspecies variations relevant to reproductive toxicity have not received much scientific attention due to a lack of prominent examples. However, there are several reports stressing the relevance of interspecies differences with regard to spermatogenesis [41] and placentation [42].

The aspect of intraspecies variations in humans has not been addressed in the current testing scheme at all. Individuals may inherit one or more genes that make it more likely to have a birth defect if the individual is exposed to certain environmental substances. Environmental substances will only induce birth defects if there is a genetic predisposition. Examples of multi-factorial birth defects include cleft palate, neural tube defect, and heart defects [43].

In order to overcome the described problems of the current testing paradigm, several regulatory framework programmes are encouraging the use of more intelligent testing designs making use of new scientific and technological achievements such as state of the art *in vitro* methods. For example the REACH legislation includes a detailed annex outlining how information requirements can be adapted in order to reduce animal testing without compromising consumer safety. The US EPA's pesticide program is also calling for a more hypothesis-driven approach to toxicology testing. The US EPA has commissioned a National Research Council (NRC) committee to provide the Agency with specific feedback and advice on critical research activity and the science and public policy issues necessary to achieve the vision. The final NRC report has recently been published [44].

## **INTERMIN STRATEGY USING REDUCTION AND REFINEMENT METHODS FOR ASSESSING REPRODUCTIVE TOXICITY**

### **The (F1-Extended) One-Generation Study**

The Agricultural Chemical Safety Assessment (ACSA) Technical Committee of the ILSI Health and Environmental Sciences Institute (HESI) has developed a testing scheme for assessing toxicity to various mammalian life stages including preconception, embryo/foetal and newborn/prewearing life stages; adolescence and adults of all ages without compromising the provision of relevant toxicological information. The focal point of the life stages paradigm is the (F1-extended) one-generation study in which mature parental males are treated for four weeks and parental females for two weeks before mating. The treatment is continued during gestation and lactation until weaning of the F1 generation offspring in order: (i) to analyse effects on fertility, (ii) to determine the effects on developing organ systems during prenatal development, and (iii) to identify any effects of developmental (gestational/lactational) exposure on the offspring. Depending on alerts observed in the onegeneration study, the test design allows continuing treatment of the parental animals, and remating them to produce a second litter. In addition, any indication for developmental neurotoxicity or developmental immunotoxicity will lead to an extension of study by integrating appropriate tests for further assessment of these effects [25]. The practical feasibility of the protocol is currently evaluated within a collaborative industrial effort with ECVAM. It should be stressed that regulatory acceptance of the protocol leading to an OECD test guideline will require a formal validation exercise fulfilling criteria as laid down in the OECD guidance document 34. The validation of the proposed study design is certainly challenging due to its complexity. In particular, the involvement of triggers requires intensive discussion when assessing the relevance and reliability of the overall protocol. However, the effort is worthwhile since the protocol is aiming to replace the two-generation study and partially also developmental toxicity testing and has therefore a saving potential of up to 50% of animals needed for reproductive toxicity testing.

### **LONG TERM AIM: INTEGRATED TESTING STRATEGIES BASED ON PREVALENCE OR ON ALERTS DERIVED FROM EXISTING KNOWLEDGE**

In order to facilitate the optimal use of alternative methods, the current paradigm that is requiring *in vivo* testing for every possible adverse outcome needs to be changed toward a hypothesis-driven strategy. Existing data on the chemical or the chemical group, *in silico* models and *in vitro* data, combined with considerations of relevance of exposure, should be used in a weight of evidence approach to waive animal intensive *in vivo* tests and should lead only in exceptional cases to confirmatory *in vivo* studies. This approach will focus on the most likely hazards of concern and determine, each chemical or chemical group and exposure scenario, what specific data are essential in order to make reliable regulatory decisions. The basis for assembling robust batteries of *in silico* and *in vitro* models will be based either on the prevalence of adverse effects (prevalence-driven strategy), or on a combination of ad-hoc testing batteries based on alerts derived from existing knowledge (alert driven strategy).

### **ALERT DRIVEN STRATEGIES**

Key parameters of integrated testing strategies are alerts deriving from other toxicological studies, observations in human populations or from the knowledge on toxicological hazards related to the chemicals structure. It is highly important that existing, non-confidential or shared toxicological data are identified and allow a detailed analysis of existing toxicological information. There is a need for guidance on how information sources such as toxicological databases, data bases for human data as well

information sources based on *in vitro* tests, QSARs, chemical categories and read-across approaches can be accessed and used.

However, in case the data are insufficient for regulatory decisions but alerts have been identified, the existing data will be the basis for the development of a tailored testing scheme. Depending on the nature of the alerts, test batteries of specific validated *in vitro* tests will be triggered in order to confirm or refute observed concerns. For example a histopathology in testes observed in repeated dose studies will be followed up by tests on spermatotoxicity models not a two-generation study, if the classification cannot be done on the finding alone. This approach is seeking for a targeted testing that provides sufficient toxicological data for hazard identification but it also keeps *in vivo* testing to a minimum. It should be stressed that *in vitro* tests should only be used as decision points in a testing strategy if validated according to internationally agreed validation criteria ensuring their reliability and relevance.

A frequent possible scenario for an alert driven strategy could be unclear histopathological observations in the testes in subacute or chronic toxicity studies. These findings should not automatically trigger additional animal intensive tests for reproductive toxicity; but these effects should be further explored by using *in vitro* testing batteries analyzing cytotoxic effects on specific cell population of the reproductive organs and/or by analyzing relevant hormone production or by monitoring gametogenesis *in vitro*. The obtained data will identify if observed changes in the tissues of reproductive organs are reprotoxic effects or if the observed effects are related to general toxicity. The establishment of relevant databases such as the one of the Fraunhofer society [34] will support the development of such a scientific approach. A query of the database containing 329 chemicals tested in repeated dose studies (rats) and 203 chemicals (mice) has demonstrated that major targets of chemicals showing toxicological effects on the testes are target cells that can also be cultured *in vitro* (Table 2). However, substantial research efforts are still necessary to maintain the functionality of target cells *in vitro* and to convert these *in vitro* models into predictive tests using specific functions as toxicological endpoints. Changes of the functionality of certain target cells will point to the relevant target mechanisms and will support the interpretation if the observed effects are relevant to humans.

## **PREVALENCE DRIVEN STRATEGY**

The prevalence driven strategy will be based on the quantification of frequency of observed adverse effects on toxicological targets of reproductive toxicants tested *in vivo*. This strategy should be applied if no information of the chemical or chemical class is available. The aspect of prevalence is highly important when setting-up testing batteries for screening purposes in order to replace *in vivo* screening tests [45]. In order to set-up a reliable database on the prevalences of subendpoints detailed analyses of existing *in vivo* reproductive toxicity studies is required. First steps leading towards the development of reliable testing strategies have been performed. Candidate chemicals with a broad database have been identified and evaluated by assessing the following databases:

1. Annex 1 of Directive 67/548/EEC. The most recent listing under Annex 1 of the EU Dangerous Substances Directive (dated 16 June 2005) was obtained from the website of the European Chemicals Bureau (ECB) and examined to identify all substances classified with the risk phrases R60 through R64.
2. The Proposition 65 list. The most recent version of the US Proposition 65 list (dated 9 June 2006) was obtained from the website of the State of California Office of Environmental Health Hazard Assessment and was likewise queried to identify all substances and/or mixtures classified as “developmental” and “male/female” reproductive toxicants. All hazard assessment documents

listed on the Proposition 65 website [46] were downloaded and reviewed individually, of which eight were found to contain relevant information for this analysis.

3. The European Chemical Substances Information System and IUCLID Chemical Data Sheet Information System were searched using individual CAS numbers as well as common chemical names. IUCLID datasets were identified and downloaded for 74 substances, of which 23 (31%) contained data on reproductive and/or developmental toxicity endpoints (i) in English and (ii) of sufficient quality to satisfy the reliability criteria established *a priori* for this analysis [47].
4. The US National Toxicology Program (NTP) [48] database of reproductive and developmental toxicity study abstracts was searched by common chemical name, which identified 43 high quality study summaries, all of which were considered to be relevant and reliable for the purposes of this analysis. A subsequent review of publications by the NTP Center for the Evaluation of Risks to Human Reproduction identified an additional eight monographs for use in this analysis.
5. The US EPA's Integrated Risk Information System [49] database was searched by CAS number and common chemical, yielding results for 17 substances. Although the IRIS reporting format was found to be overly superficial in areas normally needed to establish the reliability of a study [47], the experience with US EPA hazard and risk assessment procedures and criteria for listing in the IRIS database is sufficiently rigorous to permit use of these data without acquiring and reviewing original references.
6. Peer-reviewed open literature by analyzing studies conducted according to recognized test guidelines or which have at least examined a relatively broad range of reproductive or developmental parameters (US National Library of Medicine's TOXNET-Developmental and Reproductive Toxicity [50]).

Despite a number of exhaustive database and literature searches, data satisfying the inclusion criteria for this analysis could not be located in the public domain for more than half (53%) of the substances classified by regulators as being toxic to reproduction. Consequently, the following analysis is limited to data on 71 classified reproductive toxicants or approximately 47% of the original database. The frequencies with which statistically and biologically significant positive effects have been reported have been determined. These frequency counts are presented in Fig. (5), as counts of absolute frequency (i.e., the total number of times a positive effect was detected in a particular sub-endpoint, irrespective of the dose at which the effect was seen). The reported effects are mainly not isolated effects but they also appear in combinations. It is therefore highly relevant for a further analysis in particular for sub-endpoints occurring with a lower prevalence in order to determine if they are associated with a more frequently occurring effect. These would further diminish the relevance to test for a sub-endpoint with a lower prevalence such as e.g. parturition. Even if parturition is a sub-endpoint with a low prevalence of a health effect which has per se a low prevalence in the universe of industrial chemicals, the competent authorities are currently requesting to test for such an endpoint.

For developmental toxicity the search strategy described above identified reliable data for 202 (69%) of the classified substances. Given the extensive range of histopathological, functional, clinical and other evaluations undertaken in the context of a developmental toxicity study, standardization is important not only in relation to the selection of study endpoints, but also in the terminology used to communicate study results. For the purposes of this analysis, studies were analyzed and catalogued in a manner consistent with the recommendations of Chahoud and colleagues [51] using sub-endpoint definitions proposed by MacKenzie and Hoar [52]. The frequency with which standardized sub-endpoints from guideline prenatal developmental toxicity and developmental toxicity studies were reported positive for the 202 substances in this database is presented in Fig. (6).

### Sub-endpoints for Reproductive toxicity

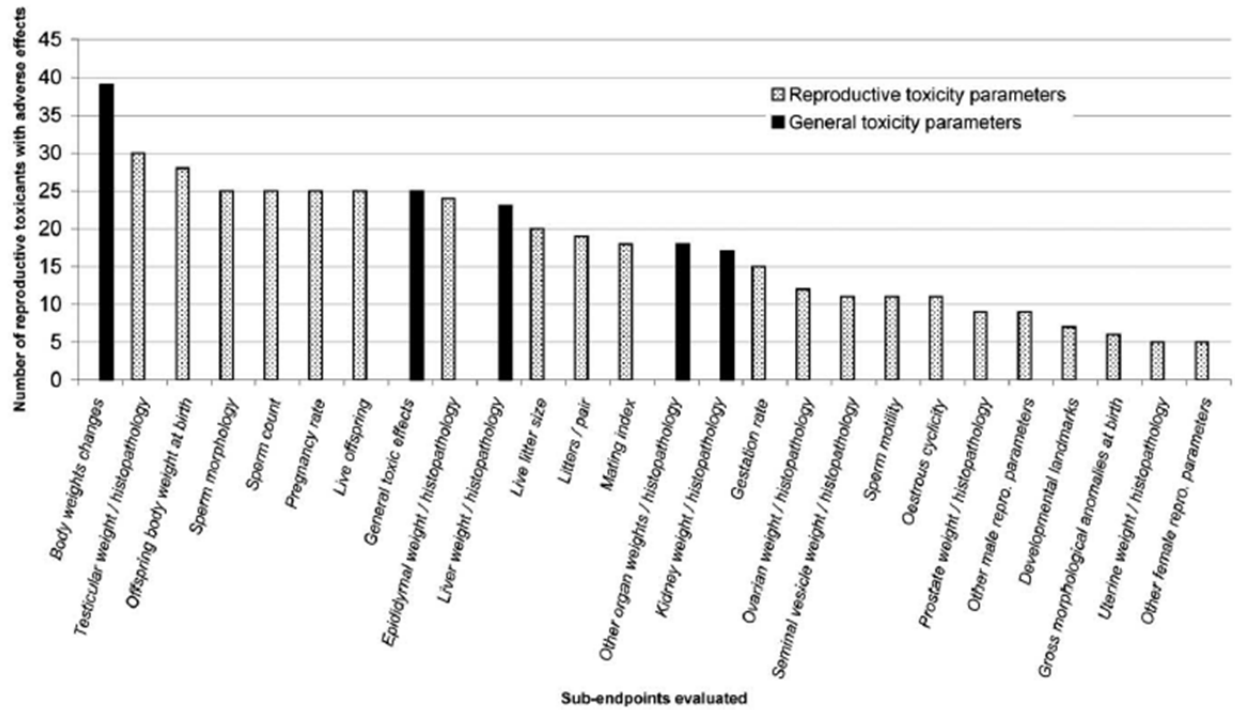


Fig. (5). Frequency with which standardised sub-endpoints yielded statistically significant findings among 71 substances classified as reproductive toxicants.

### Sub-endpoints for developmental toxicity

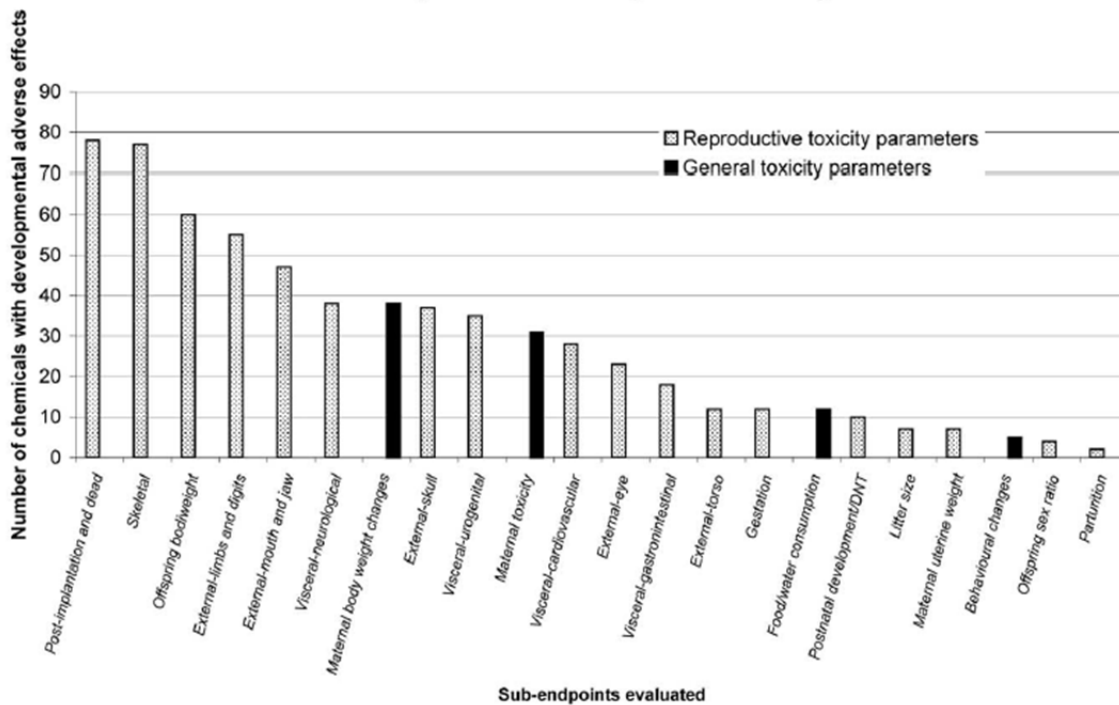


Fig. (6). Frequency of standardised sub-endpoints yielding statistically significant findings for 202 substances classified as developmental toxicants.

**Table 2. Observed Effects on the Testis**

Observed Effect	Observations Attributed to this Effect	Other Effects Observed in Combination at *Probably* Other Doses
aspermato genesis		
atrophy	seminiferous tubular atrophy; degeneration of germ cells in the epididymis or seminal duct	degeneration
cell enlargement	giant cell formation	
cell proliferation	Leydig's cell increase	
changes in organ structure	reduction in size	
degeneration	lesion of seminiferous epithelium/ -tubules; histopathological changes in Sertoli cells; atrophy; weight decreased; weight increased	changes in organ structure (decreased size and abnormal morphology); spermatogenesis impaired (including aspermia)
haemorrhage		
hypoplasia	interstitial-cell hyperplasia	atrophy
infertile testes	effect reversible	
mineralization		
necrosis	seminiferous tubules	
oedema		
spermatogenesis impaired	decreased percentage of mobile sperm; decreased sperm velocity; increase in % of abnormal sperms	weight decreased
tumour	Leydig cell tumor; malign. mesothelioma; interstitial cell adenoma	atrophy, mineralization of seminiferous tubules, mesothelioma of scrotum, weight decreased, flaccid testes (not reversible)
weight decreased	absolute and/or relative wt.	degeneration of seminiferous tubules, atrophy, vacuolization of Sertoli cells
weight increased	Relative and/or absolute wt.	

Description of observed toxicological effects if the LOEL has been established due to effects on the testis.

## **DEVELOPMENT OF A TOOL-BOX FOR ANALYZING REPRODUCTIVE TOXICITY BY USING ALTERNATIVE METHODS**

### **ReProTect**

Within the 6th Framework Programme of the European Union a international Integrated Project (ReProTect, <http://www.reprotect.eu/>), involving 32 partners from industry, academia and governmental institutions, has been set up aiming to develop and optimize *in vitro* tests that are able to detect toxic effects and mechanisms associated with reproductive toxicity. Due to the complexity of the mammalian reproductive cycle, it is not possible to model the whole cycle in one *in vitro* system in order to detect

**Table 3. Summary of Currently Pursued Tests in the Optimization Process of Test Development**

<i>In Vitro Model</i>	ECVAM Modular Approach							
	Module 1				Module 2	Module 3		Module 4
	test definition				within-laboratory variability	transferability		between-laboratory variability
	Explanation of mechanistic basis including the definition of predicted endpoints	SOP	Training set of chemicals	Definition of prediction model and domain of applicability	From 5-30 chemicals with well known defined reproductive mechanisms have been tested	Ease of transferability	Assessment of reproducibility of experimental data in second laboratory	Assessment of reproducibility of experimental data in 2-4 laboratories (done by participants not by an independent statistician)
<b>Methods for assessing mammalian fertility</b>								
CASA	Y	Y	Y	O	Y	Y	Y	Y
SCSA/Comet assay	Y	O	O	O	N	X	X	X
Leydig cells and steroidogenesis	Y	Y	Y	O	Y	X	X	X
Sertoli cells	Y	Y	Y	Y	Y	Y	Y	O
<i>In vitro</i> bovine oocyte (assessing effects on: oocyte maturation, fertilisation and pre-implantation embryo)	Y	Y	Y	Y	Y	X	X	X
<b>Methods for assessing teratogenicity</b>								
Embryonic stem cells (skeletal differentiation)	Y	Y	O	N	N	N	N	N
Embryonic stem cells (neuronal differentiation)	Y	Y	Y	O	O	N	N	N
Human embryonic stem cells test	Y	Y	N	O	N	X	X	X
<b>Methods for assessing endocrine modulation</b>								
Transcriptional tests for assessing estrogenic and androgenic activity of substances	Y	Y	Y	O	O	Y	Y	O
Receptor binding tests for estrogenic and androgenic compounds	Y	Y	Y	O	O	Y	Y	O

Y= already achieved within ReProTect N= not yet achieved O= ongoing activity.

adverse effects to mammalian reproduction. Individual tests should be used as building blocks to compose testing batteries and strategies. Within the last two years the Integrated Project ReProTect explored the predictive power of a range of pioneering *in vitro* tests, addressing to mimic key events of adverse effects with a high prevalence.

In order to speed up the process of validation of *in vitro* tests and to safeguard the consistency with internationally agreed validation criteria, the test development is following the requirements laid down in the modular approach to the ECVAM principles on test validity [36]. It is envisaged that each test will provide information on the mechanistic basis, a standard operation procedure and an assessment of the intralaboratory reproducibility of the test. A training set of chemicals will be used to define a preliminary prediction model. The respective data will be compiled in toxicological dossiers, which will be peer-reviewed before further formal validation steps can be initiated. More than 150 independently selected reproductive toxicants with different toxicological mechanisms have been selected in order to support the optimization process of test protocol development.

Adverse effects on mammalian fertility are assessed in tests reflecting various toxicological mechanisms such as effects on Leydig/Sertoli cells, folliculogenesis, germ cell maturation, the motility of sperm cells, steroidogenesis, fertilisation and on the preimplantation embryo. Another key aspect of ReProTect is the detection of teratogenic compounds. Emphasis is laid on potential of human and murine embryonic stem cells to differentiate into major target tissues of teratogens such as neural and cardiac cells as well as on cells contributing to the skeletal system. The identification of stable toxicological endpoints is a major aim of the project. Due to the high relevance of inter-species variations selected compounds will be compared in the murine and the human embryonic stem cell test. Other tests assessing (anti)-estrogenic and (anti)-androgenic activity of compounds have been optimised and are now being analysed for their predictive capacity and will so contribute to testing strategies relevant for fertility and developmental toxicity. In addition, a protocol for evaluating the ability of chemicals to bind to the estrogen receptor will now be validated under the umbrella of the respective OECD (Organization for Economic Co-operation and Development) validation management group. A summary on the current status of the tests is given in (Table 3) and will be regularly updated.

## **IN SILICO MODELS**

The development of reliable QSARs for reproductive toxicity is currently suffering due to a lack of high quality *in vivo* data and of the complexity of the reproductive toxicity endpoint involving several known and unknown toxicological mechanisms. It should be stressed that QSARs can be based either on *in vivo* or on *in vitro* data. The uncertainty of the origin of data should be taken into account when integrating these models into testing strategies. For example several QSARs that claim to predict developmental toxicity of various chemical classes are based on *Xenopus* embryos [53-55] or on *Hydra Attenuata* [56-58]. The relevance of these species for predictive human health effects needs to be carefully considered since e.g. *Hydra attenuata* is a micro-invertebrate and does not contain a third germ layer. Chemical effects on mesoderm formation and mesodermal cell differentiation can thus per se not be detected. The predictive capacity of *Xenopus* and other nonmammalian species for assessing adverse effects on human development in large scale programmes is scientifically still under discussion. Other QSARs are based on *in vitro* tests such as the whole embryo culture or primary cell cultures. The datasets of these models needs to be carefully reviewed as to their validation status before relying on predictions of the *in silico* models. The same holds true for the more promising area of QSARs estimating the receptor binding activity of estrogenic compounds [59-81]. However, the reliability of data derived from non-validated *in vitro* models feeding into the development of QSARs should be considered. Currently no *in vitro* model based on receptor binding studies has been validated according to internationally agreed criteria.



Nevertheless several *in vitro* models are currently undergoing validation exercises in order to proof their reliability and relevance for the given purpose. Since several reproductive toxicants are acting via various receptors the area of receptor mediated toxicity seems currently to be the most promising area for a successful application of QSARs if data of validated *in vitro* models are available.

Some commercially available toxicity prediction software packages are claiming to detect reproductive toxicants. Maslankiewicz *et al.* [82] have reported that the software program DEREKfW has been challenged with the around 100 reproductive toxicants included in Annex I of Directive 67/548/EEC. 90% of chemicals classified for "impaired fertility" and 81% of chemicals that cause harm to the unborn child were not detected. The TSCA chemical category list of the new chemical program of the US-EPA failed in 77% to detect EU classified chemicals causing adverse effects to mammalian fertility and 82% of developmental toxicants have not been correctly identified.

## **CONCLUSIONS**

The concept currently proposed by competent authorities for testing for every possible adverse effect of reproductive toxicants needs revision if large scale toxicological programmes should be feasible and successful. This precautionary approach will otherwise result in an enormous number of false-positive classifications. It is more advisable to develop new concepts based on the knowledge of already existing data of reproductive toxicants rather than requesting more testing following the traditional scheme. A major component of more science based concepts is a careful determination of prevalences with the support of high quality databases. Testing for the detection of adverse effects with low prevalences will lead to unnecessary animal testing but also to a high rate of false-positives.

In addition, it is necessary to increase the confidence of decision making bodies into the validation process of alternative methods. Positive results of a properly validated alternative test with a high specificity are pointing to serious hazard of the observed toxicological pathway. The added value of confirmatory *in vivo* studies should be carefully considered before just overruling a positive *in vitro* result since the observed toxicological effect can be invisible *in vivo* due to species differences e.g. ADME effects of the used laboratory animal. This does not necessarily mean that the hazard is not relevant to humans especially in different exposure scenarios. In order to support this statement the prominent example of thalidomide should be stressed again. Differentiating stem cells showed a positive signal *in vitro* when treated with thalidomide [83] in contrast to the standard rodent study which has provided an equivocal result that has been wrongly interpreted.

## **ACKNOWLEDGEMENT**

The authors would like to thank S. Eisenreich, B. Sokull-Kluttgen, T. Cole and J. Baraibar Fentanes for supporting the data analysis in the chemical database of the ECB. We would also like to thank A. Bitsch and I. Mangelsdorf for their data analysis in the databases on repeated dose studies.

## ABBREVIATIONS

ACSA	=	The Agricultural Chemical Safety Assessment
CASA	=	Computer assisted Sperm Analysis
DEREKfW	=	Deductive Estimation of Risk from Existing Knowledge for Windows
ECB	=	European Chemical Bureau
IRIS	=	Integrated Risk Information System
IUCLID	=	International Uniform Chemical Information Database
LOAEL	=	Lowest Observed Adverse Effect Level
NAS	=	National Academy of Sciences
NTP	=	National Toxicology Program
OECD	=	Organization for Economic Co-operation and Development
QSARs	=	Quantitative Structure-Activity Relationships
REACH	=	Registration, Evaluation and Authorization of Chemicals
SCSA	=	Sperm Chromatin Structure Assay
TSCA	=	Toxic Substance Control Act
US EPA	=	US Environmental Protection Agency

## REFERENCES

References 85-87 are related articles recently published in Current Pharmaceutical Design.

1. Worth AP, Balls M. Alternative (non-animal) methods for chemicals testing: current status and future prospects. *ATLA* 2002; 30 (Suppl 1): 95-102.
2. Hoffmann S, Hartung T. Toward an evidence-based toxicology. *Hum Exp Toxicol* 2006; 30: 497-513.
3. Bailey J, Knight A, Balcombe J. The future of teratology research is *in vitro*. *Biogenic Amines* 2005; 19: 97-145.
4. Pedersen F, De Bruijn J, Munn SJ, Van Leeuwen K. Assessment of additional testing needs under REACH. Effects of (Q)SARs, risk based testing and voluntary industry initiatives. Publication: EUR 20863 EN 2003.
5. van der Jagt K, Munn S, Tørsløv J, de Bruijn J. Alternative approaches can reduce the use of test animals under REACH. Addendum to the report: Assessment of additional testing needs under REACH Effects of (Q)SARs, risk based testing and voluntary industry initiatives. Publication: EUR 21405 EN 2004.
6. Schaefer C, Spielmann H, Vetter K. *Arzneiverordnung in Schwangerschaft und Stillzeit*. Jena Urban and Fischer 2006.
7. Russel WMS and Burch RL. *The Principles of Humane Experimental Technique* Methuen, London, 1959.
8. Reuter U, Heinrich-Hirsch B, Hellwig J, Holzum B, Welsch F. Evaluation of OECD screening tests 421 (reproduction/developmental toxicity screening test) and 422 (combined repeated dose toxicity study with the reproduction/developmental toxicity screening test). *Reg Toxicol Pharm* 2003; 38: 17-26.

9. Gelbke HP, Fleig H, Meder M on behalf of German Chemical Industry Association SIDS reprotoxicity screening test update: testing strategy and use. *Reg Toxicol Pharm* 2004; 39: 81-6.
10. Hoefler T, Gerner I, Gundert-Remy U, Liebsch M, Schulte A, Spielmann H, *et al.* Animal testing and alternative approaches for the human health risk assessment under the proposed new European chemicals regulation. *Arch Toxicol* 2004; 78: 549-64.
11. ECB [European Chemicals Bureau]. EU standardised testing methods adopted under annex V of council directive 67/548/EEC – Part B: Methods for the determination of toxicity. Accessed 28 July 2006 <<http://ecb.jrc.it/testing-methods>>.
12. OECD [Organisation for Economic Co-operation and Development]. OECD test guideline 415 – One-generation reproduction toxicity study. Paris (France): OECD 1983a; 8 p.
13. OECD [Organisation for Economic Co-operation and Development]. OECD test guideline 416 – Two-generation reproduction toxicity study. Paris (France): OECD 1983b; 8 p.
14. EPA [US Environmental Protection Agency]. Health effects test guidelines – OPPTS 870.3800 – Reproduction and fertility effects. August 1998(a). [http://www.epa.gov/opptsfrs/publications/OPPTS\\_Harmonized/870\\_Health\\_Effects\\_Test\\_Guidelines/Series/870-3800.pdf](http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-3800.pdf).
15. OECD [Organisation for Economic Co-operation and Development]. OECD test guideline 421 – Reproduction/developmental toxicity screening test. Paris (France): OECD 1995; 10 p.
16. EPA [US Environmental Protection Agency]. Health Effects Test Guidelines – OPPTS 870.3550 – Reproduction/developmental toxicity screening test. August 1998(c). <[http://www.epa.gov/opptsfrs/publications/OPPTS\\_Harmonized/870\\_Health\\_Effects\\_Test\\_Guidelines/Series/870-3550.pdf](http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-3550.pdf)>.
17. OECD [Organisation for Economic Co-operation and Development]. OECD test guideline 422 – Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test. Paris (France): OECD 1996; 14 p.
18. EPA [US Environmental Protection Agency]. Health effects test guidelines – OPPTS 870.3650 – Combined repeated dose toxicity study with the reproduction/developmental toxicity screening test. August 1998(d). <[http://www.epa.gov/opptsfrs/publications/OPPTS\\_Harmonized/870\\_Health\\_Effects\\_Test\\_Guidelines/Series/870-3650.pdf](http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-3650.pdf)>.
19. ICH [International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use]. Guideline S5(R2) – Detection of toxicity to reproduction for medicinal products & toxicity to male fertility. November 2005 <<http://www.ich.org/LOB/media/MEDIA498.pdf>>.
20. OECD [Organisation for Economic Co-operation and Development]. OECD test guideline 414 – Teratogenicity. Paris (France): OECD 1981; 6 p.
21. EPA [US Environmental Protection Agency]. Health effects test guidelines – OPPTS 870.3700 – Prenatal developmental toxicity study. August 1998(b). <[http://www.epa.gov/opptsfrs/publications/OPPTS\\_Harmonized/870\\_Health\\_Effects\\_Test\\_Guidelines/Series/870-3700.pdf](http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-3700.pdf)>.
22. ICH [International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use]. Guideline S5(R2) – Detection of toxicity to reproduction for medicinal products & toxicity to male fertility. November 2005 <<http://www.ich.org/LOB/media/MEDIA498.pdf>>.
23. OECD [Organisation for Economic Co-operation and Development]. Draft proposal for a new test guideline 426 – Developmental neurotoxicity. Paris (France): OECD 2006; 25 p.
24. EPA [US Environmental Protection Agency]. Health effects test guidelines – OPPTS 870.6300 – Developmental neurotoxicity study. August 1998(e). <[http://www.epa.gov/opptsfrs/publications/OPPTS\\_Harmonized/870\\_Health\\_Effects\\_Test\\_Guidelines/Series/870-6300.pdf](http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/870-6300.pdf)>.
25. Cooper RL, Lamb IV JC, Barlow SM, Bentley K, Brady AM, Doerrer NG, *et al.* A tiered approach to life stages testing for agricultural chemical safety assessment. *Crit Rev Toxicol* 2006; 36: 69-98.
26. Anon (2006) RIP 3.3: Technical Guidance Document (TGD) on Information requirements on intrinsic properties of substances. <http://ecb.jrc.it/reach/rip/>. (Accessed. 14.2.2007).

27. Fleischer M. Testing costs and testing capacity according to the REACH requirements – Results of a survey of independent and corporate GLP laboratories in the EU and Switzerland 2005; [www.euregio-reserach.eu](http://www.euregio-reserach.eu).
28. Anon (2006) Briefing note on the number of animals expected to be used under REACH summary of RE-assessment performed by the JRC. <http://ihcp.jrc.cec.eu.int/docs/ecb/REACHanimalfigures.pdf> (Accessed. 14.2.2007).
29. Wilson JD. In: Wilson JD Frazer FC Ed, Embryotoxicity of drugs to man, handbook of teratology. New York Plenum Press 1977; Vol 1: 309-55.
30. Enders G. Infektionen und Impfungen in der Schwangerschaft, 2. Aufl. Muenchen: Urban und Schwarzenberg, 1991.
31. Shardein JL. Chemically induced birth defects. 4th ed. New York, Marcel Dekker Inc. 2000.
32. Rösch C, Steinbicker V. Aetiology of congenital malformations – analysis of malformation registry data compared with the Kalter & Warkany study. *Reprod Toxicol* 2003; 17: 503-4.
33. Anon 1979 EUROCAT Congenital Anomalies and Public Health <http://www.eurocat.ulster.ac.uk/whatis.html#TheObjectivesofEUROCAT>. (Accessed 14.2.2007).
34. Bitsch A, Jacobi S, Melber C, Wahnschaffe U, Simetska N, Mangelsdorf I. REPDOSE: A database on repeated dose toxicity studies of commercial chemicals - A multifunctional tool. *Regul toxicol and pharmacol* 2006; 46: 202-10.
35. Worth A, Balls M. The Principles of Validation and the ECVAM Validation Process. *ATLA* 2002; 30(Suppl 2): 15-21.
36. Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, *et al.* A modular approach to the ECVAM principles on test validity. *ATLA* 2004; 32: 467-72.
37. OECD [Organisation for Economic Co-operation and Development]. OECD guidance document No. 34 on the validation and international acceptance of new or updated test methods for hazard assessment. Paris (France): OECD 2005; 96 p.
38. Schardein JL, Schwetz BA, Kenel MF. Species sensitivities and predictions of teratogenic potential. *Environ Health Perspective* 1985; 61: 55-67.
39. Peterson RE, Theobald HM, Kimmel GL. Developmental and reproductive toxicity of dioxins and related compounds: crossspecies comparisons. *Crit Rev Toxicol* 1993; 23: 283-335.
40. Hurtt ME, Cappon GD, Browning A. Proposal for a tiered approach to developmental toxicity testing for veterinary pharmaceutical products for food-producing animals. *Food Chem Toxicol* 2003; 41: 611-9.
41. Mangelsdorf I, Buschmann J. Extrapolation from results of animal studies to humans for the endpoint male fertility. Publication series from the Federal Institute for Occupational Safety and Health. Dortmund/Berlin/Dresden 2003.
42. Bremer S, Brittebo E, Dencker L, Eriksson G, Gustafson AL, Knudsen L, *et al.* *In vitro* tests for detecting chemicals affecting the implantation process the report and recommendation of ECVAM workshop 62 – Strategic Workshop of the EU ReProTect Project *ATLA* 2007; 35: 421-39.
43. Nora MD. Multifactorial inheritance hypothesis for the etiology of congenital heart diseases the genetic-environmental interaction. *Circulation* 1968; 38: 604.
44. NRC Committee on Toxicity and Assessment of Environmental Agents. Toxicity testing in the twenty-first century: A vision and a strategy. National Academies Press, 2007.
45. Hoffmann S, Hartung T. Diagnosis: toxic!--trying to apply approaches of clinical diagnostics and prevalence in toxicology considerations. *Toxicol Sci* 2005; 85: 422-8.
46. OEHHA [US State of California Office of Environmental Health Hazard Assessment]. Proposition 65: The safe drinking water and toxic enforcement act of 1986. Accessed 5 August 2006; <<http://www.oehha.ca.gov/prop65.html>.
47. Klimisch HJ, Andreae E, Tillmann U. A systematic approach for evaluating the quality of experimental and ecotoxicological data. *Regul Toxicol Pharmacol* 1997; 25: 1-5.

48. NTP [US National Toxicology Program]. NTP reproductive toxicity by continuous breeding study abstracts. Queried on 5 August 2006. <<http://ntp-server.niehs.nih.gov/ntpweb/index.cfm?objectid=72015F74-BDB7-CEBA-F457FDD4713C8FFD>.
49. EPA [US Environmental Protection Agency]. Integrated risk information system (IRIS) Database. Queried 15 August 2006; <<http://www.epa.gov/iris>.
50. NLM [US National Library of Medicine]. TOXNET Developmental and Reproductive Toxicology Database (DART). Queried 17 August 2006; <<http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?DARTETIC>.
51. Chahoud I, Buschmann J, Clark R, Druga A, Falke H, Faqi A, *et al.* Classification terms in developmental toxicology: need for harmonisation. *Reprod Toxicol* 1999; 13: 77-82.
52. MacKenzie KM, Hoar RM. In: Derelanko MJ, Hollinger MA Eds, *Handbook of Toxicology 2nd Ed, Developmental toxicology*. Washington (DC): CRC Press. 2002; 497-543.
53. Dawson DA. Additive incidence of developmental malformation for *Xenopus* embryos exposed to a mixture of 10 aliphatic carboxylic-acids. *Teratology* 1991; 44: 531-46.
54. Dawson DA, Schultz TW, Hunter RS. Developmental toxicity of carboxylic acids to *Xenopus* embryos: A quantitative structureactivity relationship and computer-automated structure evaluation. *Teratog Carcinog Mutagen* 1996; 16: 109-24.
55. Mekenyan OG, Schultz TW, Veith GD, Kamenska V. 'Dynamic' QSAR for semicarbazide-induced mortality in frog embryos. *J Appl Toxicol* 1996; 16: 355-63.
56. Schultz TW, Dawson DA. In: Karcher W, Devillers J Ed, *Practical applications of Quantitative Structure-Activity Relationships (QSAR) in environmental chemistry and toxicity structure-activity relationships for teratogenicity and developmental toxicity*. European Commission, Brussels 1990; 389-409.
57. Devillers J, Chezeau A, Thybaud E. PLS-QSAR of the adult and developmental toxicity of chemicals to *Hydra attenuate*. *SAR QSAR Environ Res* 2002; 13: 705-12.
58. Devillers J, Chezeau A, Thybaud E, Rahmani R. QSAR modeling of the adult and developmental toxicity of glycols, glycol ethers and xylenes to *Hydra attenuate*. *SAR QSAR Environ Res* 2002; 13: 555-66.
59. Selassie CD, Garg R, Mekapati S. Mechanism-based QSAR approach to the study of the toxicity of endocrine active substances. *Pure Appl Chem* 2003; 75: 2363-73.
60. Shi LM, Fang H, Tong WD, Wu J, Perkins R, Blair RM, *et al.* QSAR models using a large diverse set of estrogens. *J Chem Inf Comput Sci* 2001; 41: 186-95.
61. Singh AK. Development of QSAR models to predict estrogenic, carcinogenic, and cancer protective effects of phytoestrogens. *Cancer Invest* 2001; 19: 201-16.
62. Tong W, Fang H, Hong H, Xie Q, Perkins R, Sheehan DM. In: Cronin MTD, Livingstone DJ Ed, 2004. *Receptor-mediated toxicity: QSARs for estrogen receptor binding and priority setting of potential estrogenic endocrine disruptors. Predicting Chemical Fate and Toxicity*. CRC Press, Boca Raton FL 2004; 285-314.
63. Tong W, Lowis DR, Perkins R, Chen Y, Welsh WJ, Goddette DW, *et al.* Evaluation of quantitative structure-activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. *J Chem Inf Comput Sci* 1998; 38: 669-77.
64. Tong WD, Fang H, Hong HX, Xie Q, Perkins R, Anson J, *et al.* Regulatory application of SAR/QSAR for priority setting of endocrine disruptors: A perspective. *Pure Appl Chem* 2003; 75: 2375-88.
65. Tong WD, Perkins R, Strelitz R, Collantes ER, Keenan S, Welsh WJ, *et al.* Quantitative structure-activity relationships (QSARs) for estrogen binding to the estrogen receptor: Predictions across species. *Environ Health Perspect* 1997; 105: 1116-24.
66. Tong WD, Xie W, Hong HX, Shi LM, Fang H, Perkins R. Assessment of prediction confidence and domain extrapolation of two structure-activity relationship models for predicting estrogen receptor binding activity. *Environ Health Perspect* 2004; 112: 1249-54.
67. Waller CL, Minor DL, McKickinney JD. Using 3-dimensional quantitative structure-activity-relationships to examine estrogenreceptor binding affinities of polychlorinated hydroxybiphenyls. *Environ Health Perspect* 1995 103: 702-7.
68. Waller, CL. A comparative QSAR study using CoMFA, HQSAR, and FRED/SKEYS paradigms for estrogen receptor binding affinities of structurally diverse compounds. *J Chem Inf Comput Sci* 2004; 44: 758-65.

69. Waller CL, Oprea TI, Chae K, Park HK, Korach KS, Laws SC, *et al.* Ligand-based identification of environmental estrogens. *Chem Res Toxicol* 1995; 9: 1240-8.
70. Wolohan P, Reichert DE. CoMFA and docking study of novel estrogen receptor subtype selective ligands. *J Comput Aided Mol Des* 2003; 17: 313-28.
71. Xing L, Welsh WJ, Tong W, Perkins R, Sheehan DM. Comparison of estrogen receptor alpha and beta subtypes based on comparative molecular field analysis (CoMFA). *SAR QSAR Environ Res* 1999; 10: 215-37.
72. Yu SJ, Keenan SM, Tong W, Welsh WJ. Influence of the structural diversity of data sets on the statistical quality of three-dimensional quantitative-activity relationship (3D-QSAR) models: Predicting the estrogenic activity of xenoestrogens. *Chem Res Toxicol* 2002; 15: 1229-34.
73. Schmieder P, Mekenyan O, Bradbury S, Veith, G. QSAR prioritization of chemical inventories for endocrine disruptor testing. *Pure Appl Chem* 2003; 75: 2389-96.
74. Schmieder PK, Ankley G, Mekenyan O, Walker JD, Bradbury S. Quantitative structure-activity relationship models for prediction of estrogen receptor binding affinity of structurally diverse chemicals. *Environ Toxicol Chem* 2003; 22: 1844-54.
75. Saliner AG, Netzeva TI, Worth AP. Prediction of estrogenicity: validation of a classification model. *SAR QSAR Environ Res* 2006; 17: 195-223.
76. Saliner AG, Amat L, Carbo-Dorca R, Schultz TW, Cronin MTD. Molecular quantum similarity analysis of estrogenic activity. *J Chem Inf Comput Sci* 2003; 43: 1166-76.
77. Netzeva TI, Gallegos-Saliner A, Worth AP. Comparison of the applicability domain of a quantitative structure-activity relationship for estrogenicity with a large chemical inventory. *Environ Toxicol Chem* 2006; 25: 1223-30.
78. Mukherjee S, Mukherjee A, Saha A. QSAR studies with E-State index: Predicting pharmacophore signals for estrogen receptor binding affinity of triphenylacrylonitriles. *Biol Pharm Bull* 2005; 28: 154-57.
79. Mukherjee S, Saha A, Roy K. QSAR of estrogen receptor modulators: exploring selectivity requirements for ER alpha versus ER beta binding of tetrahydroisoquinoline derivatives using E-state and physicochemical parameters. *Bioorg Med Chem Lett* 2005; 15: 957-61.
80. Mekenyan O, Kamenska V, Serafimova R, Poellinger L, Brouwer A, Walker J. Development and validation of an average mammalian estrogen receptor-based QSAR model. *SAR QSAR Environ Res* 2002; 13: 579-595.
81. Kramer VJ, Giesy JP. Specific binding of hydroxylated polychlorinated biphenyl metabolites and other substances to bovine calf uterine estrogen receptor: structure-binding relationships. *Sci Total Environ* 1999; 233: 141-161.
82. Maslankiewicz L, Hulzebos EM, Vermeire TG, Müller JJA, Piersma AH. Can chemical structure predict reproductive toxicity? (2005) RIVM report 601200005/ <http://rivm.openrepository.com/rivm/bitstream/10029/7374/1/601200005.pdf> (Accessed 15.2.2007).
83. zur Nieden NI, Kempka G, Ahr HJ. Molecular multiple endpoint embryonic stem cell test- a possible approach to test for the teratogenic potential of compounds. *Toxicol Appl Pharmacol* 2004; 194, 257-69.
84. Kimmel K, Benson B, Foureman G, *et al.* A review of the reference dose and reference concentration processes. December 2002. <<http://www.epa.gov/osa/spc/rfdproc.htm>>.
85. Chengalvala MV, Meade EH Jr, Cottom JE, Hoffman WH, Shanno LK, Wu MM, *et al.* Regulation of female fertility and identification of future contraceptive targets. *Curr Pharm Des* 2006; 12(30): 3915-28.
86. Collins TF. History and evolution of reproductive and developmental toxicology guidelines. *Curr Pharm Des* 2006; 12(12): 1449-65.
87. Flick B, Klug S. Whole embryo culture: an important tool in developmental toxicology today. *Curr Pharm Des* 2006; 12(12): 1467-88.