OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters as a basis for regulation of substances with endocrine disrupting properties

Ulla Hass, Majken Dalgaard, Kirsten Jarfelt and Thuri S.A. Kledal

Department of Toxicology and Risk Assessment,
Danish Institute for Food and Veterinary Research, Denmark

TemaNord 2004:555

OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters as a basis for regulation of substances with endocrine disrupting properties

TemaNord 2004:555

ISBN 92-893-1073-1
ISSN 0908-6692

Nordic Council of Ministers
Store Strandstræde 18
DK-1255 Copenhagen K
Phone (+45) 3396 0200
Fax (+45) 3396 0202

Nordic Council
Store Strandstræde 18
DK-1255 Copenhagen K
Phone (+45) 3396 0400
Fax (+45) 3311 1870

www.norden.org

Nordic Environmental Co-operation

Environmental co-operation is aimed at contributing to the improvement of the environment and forestall problems in the Nordic countries as well as on the international scene. The co-operation is conducted by the Nordic Committee of Senior Officials for Environmental Affairs. The co-operation endeavours to advance joint aims for Action Plans and joint projects, exchange of information and assistance, e.g. to Eastern Europe, through the Nordic Environmental Finance Corporation (NEFCO).

The Nordic Council of Ministers
was established in 1971. It submits proposals on co-operation between the governments of the five Nordic countries to the Nordic Council, implements the Council's recommendations and reports on results, while directing the work carried out in the targeted areas. The Prime Ministers of the five Nordic countries assume overall responsibility for the co-operation measures, which are co-ordinated by the ministers for co-operation and the Nordic Co-operation committee. The composition of the Council of Ministers varies, depending on the nature of the issue to be treated.

The Nordic Council
was formed in 1952 to promote co-operation between the parliaments and governments of Denmark, Iceland, Norway and Sweden. Finland joined in 1955. At the sessions held by the Council, representatives from the Faroe Islands and Greenland form part of the Danish delegation, while Åland is represented on the Finnish delegation. The Council consists of 87 elected members - all of whom are members of parliament. The Nordic Council takes initiatives, acts in a consultative capacity and monitors co-operation measures. The Council operates via its institutions: the Plenary Assembly, the Presidium and standing committees.

Contents

Preface

There is a concern that exposure to endocrine disrupting chemicals (EDCs) may be harmful to humans and the environment. Therefore, it is important that EDCs will be covered properly by the chemicals legislation in force.

The existing chemicals legislation is changing these years. A new EU chemicals legislation, Registration, Evaluation and Authorisation of Chemicals (REACH), is under way as well as a Globally Harmonised System of hazard classification and labelling of chemicals (GHS). The existing chemical legislation only covers some of the effects caused by endocrine disruption. In the proposal for the new REACH system, endocrine disrupters are covered by the authorisation procedure based on a case-by-case assessment, but no indication of criteria for the assessment is given.

The OECD Task force on Endocrine Disrupter Testing and Assessment (EDTA) has agreed upon a revised Conceptual Framework (CF) for Testing and Assessment of potential endocrine disrupting substances in June 2002. Several toxicological and ecotoxicological screening tests for predicting endocrine disrupting properties of a chemical are now under development and international validation, but it will probably take several years before the full range of validated testing methods and criteria are developed.

In the meantime there is a need for guidance on how to interpret test results from existing test methods and how to identify a substance as an endocrine disrupter. Regulatory instruments that can be used towards endocrine disrupters should also be considered in this interim period.

Therefore, the Nordic co-ordination group for test method development (Nord-Utte) has initiated a project with the aim to investigate if and how the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupters can be used as a basis for regulatory initiatives towards endocrine disrupters. The project includes an assessment of the tests in the OECD Conceptual Framework, including specification of the endpoint for the test and reliability and relevance for effects in humans.

In general, the report is expected to serve as a basis for the Nordic contribution to the discussions in EU on interpretation and use of test results indicating endocrine disruption for regulatory purposes, and furthermore how to integrate endocrine disrupters in the new EU chemicals regulation.

The report has been prepared by a working group from Department of Toxicology and Risk Assessment at the Danish Institute for Food and Veterinary Research (Institute of Food Safety and Nutrition, Danish Veterinary and Food Administration until 31 December 2003).

A project group, appointed by Nord-Utte, has functioned as a sparring partner for the working group, by giving input to the work, participating in discussions and commenting on drafts during the process. Members of the project group were:

Pia Juul Nielsen, Danish Environmental Protection Agency (chairman)

Anneke Frøysa, Norwegian Pollution Control Authority

Christine Bjørge, Norwegian Institute of Public Health

Agneta Ohlsson, National Chemicals Inspectorate, Sweden

Anneli Kojo, National Product Control Agency for Welfare and Health, Finland

This work has been funded by Nord-Utte, under the Nordic Chemicals Group, on behalf of the Nordic Council of Ministers.

April 2004

List of abbreviations

AGD	Anogenital distance
AR	Androgen receptor
BPA	Bisphenol A
bw	Body weight
CALUX	Chemical Activated LUciferase gene eXpression
CHO	Chinese hamster ovary cells
CMRs	Carcinogens, or mutagens or substances that are toxic to reproduction
CNS	Central nervous system
DEHP	Di(2-ethylhexyl)phthalate
DES	Diethylstilbestrol
DHT	Dihydrotestosterone
DRP	Detailed Review Paper
E2 and E	Oestrogen
ECT	Endocrine challenge test
ED	Endocrine disrupter
EDC	Endocrine disrupting chemical
EDSTAC	Endocrine Disrupter Screening and Testing Advisory Committee
EDTA	Endocrine disrupters testing and assessment
EPA	Environmental Protection Agency
ER	Oestrogen receptor
ERE-LUC	Oestrogen receptor response-element-luciferase vector
EROD	Ethoxy-resorufin dealkylase (microsomal phase I enzyme)
EU	European Union
EU-RAR	EU-risk assessment report
FSH	Follicular stimulating hormone
GD	Gestation day
GHS	Globally Harmonised System
GnRH	Gonadotrophin releasing hormone
hCG	Human chorion gonadotrophin
hER	Human estrogen receptor
HPA	Hypothalamus pituitary axis
HPG	Hypothalamus pituitary gonadal
I.p.	Intra peritoneal
IPCS	International Programme on Chemical Safety
Lat	Lateral
LE	Long Evans
LH	Lutenizing hormone
LOAEL	Lowest Observed Adverse Effect Level
MCF-7	Michigan Cancer Foundation human mammary carcinoma cell line
MMTV-LUC	Mouse mammary tumor virus-luciferase
mRNA	Messenger ribo-nucleic acid
MVLN	A MCF-7 derivative containing an ER-controlled segment of the vitellogenin promotor

MW	Molecular weight
NOAEL	No observed adverse effect level
NP	Nonylphenol
OECD	Organisation for Economic Development and Cooperation
OP	Octylphenol
P	Progesterone
P.O.	Peroral / oral dosing by gavage
p.p-DDE	The predominant component of DDT (Synonyms for p,p'-DDE*:* DDE; 1,1'-(2,2-dichloroethenylidine) bis(4-chloro)benzene; benzene, 1,1'-(dichloroethylidene)bis(4-chloro- (9CI); 1,1-dichloro-2,2-bis(p-chlorophenyl)-ethylene; p,p'-dichlorodiphenyldichloroethylene; p,p'-dichlorodiphenoldichloroethylene
PBDE	Polybrominated diphenyl ether (flame retardant)
PCB	Polychlorinated biphenyl
PND	Postnatal day
PNS	Peripheral nervous system
PPS	Preputial separation
PRL	Prolactin
QSAR	Quantitative structure-activity relationship
REACH	Registration, evaluation and authorisation of chemicals
s.c.	Subcutaneous
SD	Sprague Dawley
StAR	Steroid acute regulatory protein
Sv	Seminal vesicles
T	Testosterone
T3	Triiiodothyronine
T4	Thyroxine
TG	Test guidelines
TGD	Technical Guidance Document
TH	Thyroid hormone
TR	Thyroid receptor
TSH	Thyroid stimulating hormone
TTR	TransThyRetin
VO	Vaginal opening
Vp	Ventral prostate
VS	Vesicula seminalis
VTG	Vitellogenin
Wt	Weight
WWF	World Wildlife Foundation

Executive Summary

Background and aim of the report

There is a growing concern on possible harmful consequences of exposure to chemicals that are capable of modulating or disrupting the endocrine system. The Nordic countries are also concerned that exposure to endocrine disrupting chemicals (EDCs) may be harmful to humans and the environment. Therefore, it is considered important that EDCs will be covered properly by the chemicals legislation in force.

Existing chemicals legislation is changing these years. A new EU chemicals legislation, REACH (Registration, Evaluation and Authorisation of Chemicals), is under way as well as the Globally Harmonised System of classification and labelling of chemicals (GHS).

According to the EU White Paper "Strategy for a future Chemicals Policy" endocrine disrupters should be classified in accordance with the existing chemicals legislation, which however, should be brought in line with the GHS. Several chemicals with endocrine disrupting properties are carcinogenic or toxic to reproduction or have properties warranting classification for chronic toxicity to humans according to the existing regulation. In many cases, however, it is still unclear whether these effects are caused by endocrine disrupting properties. It is also unclear if the existing classification criteria will lead to classification for all types of effects that can be seen after exposure to endocrine disrupters, e.g. thyroid effects and developmental neurotoxicity effects.

The European Commission's proposed new chemicals legislation, REACH, states that chemicals meeting certain criteria for very high concern, e.g. carcinogens or substances toxic to reproduction, should be brought into an authorisation scheme. The European Commission's White Paper states that the majority of endocrine disrupting chemicals would have to undergo authorisation. It highlights that the health effects, which have so far been associated with endocrine disrupting chemicals, would qualify a substance either to be classified as a carcinogen or as toxic to reproduction and so would trigger its submission to authorisation. This may however, be an optimistic interpretation of the scope of the current REACH proposal, because the Commission's proposal only relates to CMR (Carcinogens, Mutagens, Reproductive toxicants) placed in categories 1 and 2. This would for example exclude bisphenol A and nonylphenol from authorisation, because these chemicals are classified as toxic to reproduction in category 3.

In June 2001, the Environment Council concluded that endocrine disrupters should be covered by the authorisations procedure in REACH, when scientifically valid test methods and criteria have been established. In the Commission proposal of REACH from October 2003, endocrine disrupters may also be included in the authorisation procedure based on a case-by-case assessment.

OECD established a Task Force on Endocrine Disrupter Testing and Assessment (EDTA) in 1996 under the Test Guideline Programme. The aim was to develop methods for assessment and testing of chemicals with endocrine disrupting properties. At the 6th EDTA meeting in June 2002, a revised Conceptual Framework for Testing and Assessment of potential endocrine disrupting substances was confirmed (shown in

Figure 1 in the report). Several toxicological and ecotoxicological screening tests are under development.

It will most probably take several years before validated test methods and criteria are developed. In the meantime there is need for guidance in doing the case-by-case assessment, e.g. how to interpret test results and identify a substance as being an endocrine disrupter, as well as a need for considering regulatory initiatives in relation to EDCs.

Therefore, the Nordic co-ordination group for test method development (Nord-Utte) has initiated this project with the aim to investigate how the OECD Conceptual Framework can be used as a basis for regulatory initiatives towards endocrine disrupters.

The focus of this report is limited to human health effects, especially effects of oestrogenic and androgenic agonistic or antagonistic activity with regard to effects on human reproduction. Effects on the thyroid system are included to a limited extent.

Thus, the project includes an assessment of the tests in the OECD Conceptual Framework, including specification of the endpoint for the test and reliability and relevance for effects in humans.

The report is expected to serve as a basis for the Nordic contribution to the discussions in EU about how to interpret and use test results that indicate endocrine disruption for regulatory purposes, and furthermore how to integrate endocrine disrupters in the new EU chemicals regulation.

Conclusions

In general, it is considered important to recognise that case-by-case evaluation of EDC data requires special expertise, because of the complexity of ED effects and the rapid expansion of the knowledge on EDCs. Therefore, experts in the field should evaluate the data on potential EDCs.

Positive in vitro test results from well-performed studies indicate potential EDC activity in vivo and the mechanism of action may generally be considered relevant for humans. Therefore, reliable in vitro data can be used to place the chemical as a category 2 substance on the EU list of potential endocrine disrupting chemicals prioritised for further testing. Some examples indicate that chemicals with high potency in vitro may also have a high potency in vivo. If humans are exposed to a substantial degree to a chemical tested positive, further in vivo testing should be given a high priority. Regulatory actions until in vivo results become available might also be considered case-by-case, if the potency of a chemical in vitro is of similar magnitude as well known EDCs.

Negative in vitro test results cannot be used to exclude potential EDC activity, because of limitations such as inability or unknown capacity to metabolically activate chemicals and because EDC activity can occur through mechanism other than those tested in in vitro test systems.

QSAR models for ED activity and effects are under development, but at present the use for priority setting and hazard and risk assessment has not been decided.

The Uterotrophic and Hershberger assay presently being internationally evaluated under the OECD Test Guideline Program appear reliable in identifying substances with

(anti)oestrogenic or (anti)androgen mode of action. The mechanisms and the effects on most of the target tissues are highly relevant for humans. Positive results can be used to place the chemical as a category 1 substance on the EU list of potential endocrine disrupting chemicals prioritised for further testing. The assays provide in vivo NOEL/LOELs for the endpoints examined. For several EDCs, the dose levels causing effects in these assays seem to be of a similar magnitude or higher than those causing effects in reproductive and developmental toxicity studies such as the OECD two-generation study. As results from reproductive and developmental toxicity testing may take several years to obtain, it seems warranted to perform preliminary risk assessment of the chemicals based on the NOEL/LOELs from Uterotrophic and Hershberger assay, if available. However, the use of a larger margin of safety than that used based on data from e.g. the OECD two-generation study should be considered, because developmental toxicity effects in some cases have been shown to occur at lower dose levels than those causing effects in these short-term in vivo assays. Results from the Uterotrophic and Hershberger assay may also be useful when considering hazard classification of a chemical for e.g. reproductive and developmental toxicity.

A negative uterotrophic response, in a thorough dose-response study indicates that the test compound is not an ER-ligand in vivo. Equally a negative response in the Hershberger assay indicates that the test compound is neither an AR-ligand nor a 5-alpha reductase inhibitor in vivo. A test compound found negative in these assays may still have endocrine disrupting properties mediated through other mechanisms.

A number of assays may provide information on the ability of a substance to act on the production of steroids. Positive results indicate that the substance may cause adverse effects in a two-generation test. The ex vivo methods are used to assess substances for altering steroid production and secretion. A positive result in the ex vivo studies indicates a potential for effects in vivo and may as such give some basis for concern. Generally, the results of these tests can be used to place the chemical as a category 2 substance on the EU list of potential endocrine disrupting chemicals prioritised for further testing. In addition, results demonstrating clear effects on production of steroids, especially the testosterone surge during prenatal development of the males, may be used similarly as results from the uterotrophic or Hershberger assay for preliminary risk assessment and for hazard classification purposes.

The pubertal assays, the intact male assay, and the enhanced OECD TG 407 provide information about the potency of the compound in vivo. Effects on the various endpoints included in these assays can be considered adverse and/or as representing an effect on a mechanism relevant for humans. Therefore, these assays can be used to provide NO(A)ELs/LO(A)ELs to be used in human risk assessment. The use of a larger margin of safety than that used based on results from e.g. the OECD two-generation study should be considered, because developmental toxicity effects in several cases have been shown to occur at lower dose levels than those causing effects in these assays.

The reproductive and developmental toxicity studies provide adverse effect data and are used for risk assessment and hazard classification, as the results indicate potential for effects in humans. A number of potential enhancements of the existing guidelines in order to detect effects of EDCs seem relevant and lack of effects in the current reproductive toxicity studies such as e.g. the OECD two-generation study can therefore at present not exclude the possibility for EDC effects. The effects observed in

reproductive toxicity studies may be due to other mechanism than endocrine disruption, however, the pattern of effects may indicate that endocrine effects are involved. For example, a pattern of decreased anogenital distance, retained nipples and effects on reproductive organs in male offspring indicates that antiandrogenic effects are involved, while early sexual maturation in females in the absence of effects on body weight indicates that oestrogenic effects may be involved. In such cases, the results can be used to place the chemical on the EU list of potential endocrine disrupting chemicals.

Proposals

Based on considerations of sensitivity to EDC effects, enhancements of the existing long term OECD TG for reproductive and developmental toxicity is proposed. The need for these enhancements should be considered in relation to the testing strategy used. The proposed enhancements are:

- TG 407, repeated dose 28-day oral toxicity study in rodents (enhanced): Include spermatogenic staging in the histopathological examination of the testes, in order to compensate for not dosing throughout a complete spermatogenic cycle.
- TG 414, prenatal developmental toxicity studies: Consider including measurements of testicular testosterone levels in GD 21 foetuses.
- TG 415, one-generation study: Update to include similar endpoints as the two-generation study, e.g. sperm analysis, oestrus cyclicity and histopathology in paternal animals. In addition, consider extending the exposure period to postnatal day 90 instead of day 21 and include assessment of anogenital distance, nipple retention, sperm and oestrus cyclicity endpoints as well as histopathological investigations of reproductive organs in the offspring.
- TG 416, two-generation study: Include assessment of anogenital distance and nipple retention in F1 and F2 and investigations of malformations of the reproductive organs in more than one offspring per sex per litter.
- TG 426, developmental neurotoxicity studies: Evaluate effects on sexual dimorphic behaviour and include assessment of e.g. mating behaviour.

Concerning hazard classification, it has been suggested by WWF that EDCs should be placed in ED sub-categories depending on the level of available evidence for endocrine disruption. We find that an extended use of the available data on EDC effects could be more feasible and relevant. Therefore in relation to hazard classification for reproductive toxicity, we propose:

- Evidence from *in vitro* testing or *in vivo* screening can be used as 'other relevant information' demonstrating that the chemicals operate with a mechanism relevant for humans. This information can support upgrading from reproductive toxicity, category 3 to category 2 in cases where it is debated whether reproductive and developmental toxicity effects should be considered as adverse.
- Evidence from in vivo screening models such as e.g. the Hershberger or Uterotrophic assay can be used directly for hazard classification, because positive test results indicate reproductive and developmental toxicity in tests such as the OECD two-generation study at dose levels of similar or lower magnitude.

Consequently, it seems warranted to classify such chemicals for reproductive toxicity at least in category 3.

Upgrading from category 3 to 2 means that the chemical will be triggered for authorisation. However, the chemicals placed in category 3 based on in vivo evidence of ED activity and a strong suspicion of potential developmental toxicity will not automatically be triggered for authorisation. Based on the strong suspicion of developmental toxicity effects, we find that authorisation of such chemicals should also be considered

Sammenfatning

Baggrund og formål

Der er stigende bekymring for, at udsættelse for kemikalier, der er istand til at påvirke eller forstyrre hormonsystemet, kan have alvorlige konsekvenser. De nordiske lande er også bekymrede for at udsættelse for hormonforstyrrende stoffer kan medføre skader på mennesker og miljø. Derfor er det meget vigtigt, at hormonforstyrrende stoffer er tilstrækkeligt omfattet af den gældende kemikalielovgivning.

Kemikalielovgivningen ændres i disse år. En ny EU-kemikalieregulering REACH (Registrering, Evaluering og Autorisation af kemikalier) er på vej ligesom et globalt harmoniseret system til klassificering og mærkning af kemikalier (GHS).

Ifølge EU's hvidbog om en strategi for en fremtidig kemikaliepolitik skal hormonforstyrrende stoffer klassificeres i overensstemmelse med eksisterende kemikalielovgivning, som dog skal tilpasses, så den følger det nye globaliserede system GHS. Mange kemiske stoffer, som har hormonforstyrrende egenskaber, er kræftfremkaldende, medfører skader på reproduktionen eller har egenskaber, som kræver klassificering for kroniske effekter hos mennesker i henhold til den eksisterende regulering. Imidlertid er det i mange tilfælde stadig usikkert, om disse effekter er forårsaget af de hormonforstyrrende egenskaber. Det er også uklart, om de eksisterende klassificeringskriterier vil føre til klassificering af alle de typer af effekter, som kan påvises efter udsættelse for hormonforstyrrende stoffer, for eksempel effekter på skjoldbruskkirtlen og effekter på hjernens udvikling.

EU-Kommissionens forslag til en ny kemikalielovgivning, REACH, anfører, at kemiske stoffer, som opfylder bestemte kriterier for særligt bekymrende stoffer, eksempelvis kræftfremkaldende stoffer eller stoffer som skader reproduktionen, skal omfattes af en særlig autorisationsordning. Af EU-Kommissionens hvidbog fremgår også, at hovedparten af de hormonforstyrrende stoffer vil blive omfattet af autorisationsproceduren. Hvidbogen fremhæver, at de sundhedseffekter, som hidtil er blevet sat i forbindelse med hormonforstyrrende stoffer, i sig selv vil medføre at stofferne enten skal klassificeres som kræftfremkaldende eller skadelige for reproduktionen, og dette vil udløse et krav om autorisation. Imidlertid synes dette at være en optimistisk fortolkning af omfanget af det aktuelle forslag til REACH, idet Kommissionens forslag kun omfatter CMR-stoffer (kræftfremkaldende, mutagene og reproduktionsskadende), der er klassificeret i kategori 1 og 2. For eksempel vil bisphenol A og nonylphenol blive undtaget fra autorisation, fordi disse stoffer er klassificerede som skadelige for reproduktionen i kategori 3.

I juni 2001 konkluderede Miljørådet, at hormonforstyrrende stoffer skulle omfattes af autorisationsproceduren i REACH, når videnskabeligt accepterede testmetoder og kriterier er blevet fastlagt. I følge Kommissionens forslag til REACH fra oktober 2003 kan hormonforstyrrende stoffer også blive omfattet af autorisationsproceduren ved en case-by-case vurdering.

I 1996 etablerede OECD en "Task Force" for testning og vurdering af hormonforstyrrende stoffer (EDTA) under testguideline-programmet. Formålet var at udvikle

metoder til vurdering og testning af kemikalier med hormonforstyrrende egenskaber. Ved det sjette EDTA-møde i juni 2002 blev der opnået enighed om et revideret "conceptual framework" for testning og vurdering af mulige hormonforstyrrende stoffer (se figur 1 i rapporten). Adskillige toksikologiske og økotoksikologiske testmetoder til screening af stoffer for hormonforstyrrende egenskaber er nu under udvikling.

Det vil højst sandsynligt tage mange år før validerede testmetoder og kriterier er blevet udviklet. I den mellemliggende periode er der behov for vejledning i forbindelse med case-by-case vurderinger, eksempelvis vejledning i hvordan testresultater skal fortolkes og hvornår et stof kan identificeres som værende hormonforstyrrende.

Den nordiske gruppe for koordinering af testmetoder, Nord-Utte, har derfor igangsat dette projekt med det formål at undersøge, om og hvordan OECD's conceptual framework kan anvendes som basis for regulatoriske initiativer i forhold til hormonforstyrrende stoffer.

Fokus for denne rapport begrænser sig til effekter på menneskers sundhed, og i særlig grad til effekter af østrogen og androgen agonistisk og antagonistisk aktivitet i forhold til effekter på menneskets reproduktion. Effekter på det thyroide system er medtaget i begrænset omfang.

Projektet omfatter således en vurdering af testene i OECD's conceptual framework, herunder specifikation af endpoint for hver testmetode samt pålidelighed og relevans for effekter i mennesker.

Rapporten forventes at kunne danne basis for det nordiske bidrag til diskussioner i EU om, hvordan testresultater, som indikerer hormonforstyrrelser, skal fortolkes og bruges i regulatorisk sammenhæng, og endvidere hvordan hormonforstyrrende stoffer skal integreres i den nye EU kemikalielovgivning.

Konklusioner

Helt generelt er det vigtigt at være opmærksom på, at case-by-case vurdering af data vedrørende hormonforstyrrende effekter kræver særlig ekspertise, dels på grund af kompleksiteten af hormonforstyrrende effekter og dels på grund af den hurtige øgning af viden om hormonforstyrrende stoffer.

Positive in vitro testresultater fra velgennemførte undersøgelser indikerer mulig hormonforstyrrende aktivitet in vivo, og virkningsmekanismen kan generelt anses for at være relevant for mennesker. Derfor kan pålidelige in vitro data anvendes til at indplacere et kemisk stof som et kategori 2 stof på EU's liste over mulige hormonforstyrrende stoffer, der er prioriteret til yderligere testning. Nogle undersøgelser indikerer, at kemikalier med høj potens in vitro også har høj potens in vivo. Hvis mennesker i betydelig grad bliver udsat for et kemisk stof, som er testet positivt, bør yderligere in vivo testning prioriteres højt. Regulatoriske tiltag indtil in vivo resultater er tilgængelige må også overvejes i hvert enkelt tilfælde, hvis det kemiske stofs potens in vitro er af samme størrelsesorden som velkendte hormonforstyrrende stoffer.

Negative in vitro test resultater kan ikke anvendes til at udelukke mulig hormonforstyrrende aktivitet, grundet begrænsninger som manglende eller ukendt kapacitet for metabolisk aktivering af det kemiske stof, og fordi hormonforstyrrende

aktivitet kan optræde via andre mekanismer end dem, der undersøges for i in vitro testsystemer.

QSAR-modeller for hormonforstyrrende aktivitet og effekter er under udvikling, men på nuværende tidspunkt er anvendelsen af QSAR til prioritering og fare- og risikovurdering ikke blevet besluttet.

Uterus-testen og Hershberger-testen, som i øjeblikket er under international evaluering under OECD's testguideline program, ser ud til at være pålidelige til at påvise stoffer med østrogen/anti-østrogen og androgen/anti-androgen virkningsmåde. Mekanismerne og effekterne på de fleste målorganer/væv er meget relevante for mennesker. Positive resultater kan anvendes til at indplacere det kemiske stof som et kategori 1 stof på EU's liste over mulige hormonforstyrrende stoffer prioriteret til yderligere testning. Disse test giver in vivo NOEL/LOELs for de undersøgte endpoints. For en del hormonforstyrrende stoffer synes de dosisniveauer, som medfører effekter i disse tests, at være af samme størrelsesorden eller højere end dem, som medfører effekter i undersøgelser for reproduktions- og udviklingstoksicitet, som for eksempel OECD's to-generationsstudie. Da resultater fra reproduktions- og udviklingstoksicitets-undersøgelser kan tage en del år at fremskaffe, synes det påkrævet at gennemføre en foreløbig risikovurdering af det kemiske stof baseret på NOEL/LOELs fra Uterus-testen og Hershberger-testen, hvis de er tilgængelige. Imidlertid bør anvendelse af en større sikkerhedsmargin end den, der anvendes baseret på data fra OECD to-generationsstudiet, overvejes, fordi toksiske effekter på udviklingen i nogle tilfælde har vist sig at forekomme ved lavere dosisniveauer end dem, der medfører effekter i disse korttids in vivo tests. Resultater fra Uterus- og Hershberger-testen kan også være nyttige i forbindelse med overvejelse af fareklassificering af et kemisk stof for f.eks. reproduktions- og udviklingstoksicitet.

Et negativt respons i Uterus-testen i et grundigt dosis-respons studie indikerer, at teststoffet ikke er en ER-ligand in vivo. På samme måde indikerer et negativt respons i en Hershberger-test, at teststoffet hverken er en AR-ligand eller en 5-alfa-reduktasehæmmer in vivo. Et teststof, som er fundet negativt i disse tests kan godt have hormonforstyrrende egenskaber, som udvises via andre mekanismer.

En del tests kan give viden om et stofs evne til at påvirke produktionen af steroider. Positive resultater indikerer, at stoffet kan medføre skadelige effekter i et to-generationsstudie. Ex vivo metoderne anvendes til at undersøge, om stoffer ændrer steroidproduktion og -sekretion. Et positivt resultat i ex vivo undersøgelserne indikerer en mulighed for effekter in vivo, og må alt andet lige føre til en vis bekymring. Generelt kan resultater fra disse tests, anvendes til at indplacere et kemisk stof som et kategori 2 stof på EU's liste over mulige hormonforstyrrende stoffer prioriteret til yderligere testning. Desuden kan resultater, som viser klare effekter på produktionen af steroider, især testosteronbølgen under den hanlige fosterudvikling, anvendes på tilsvarende vis som resultater fra Uterus- og Hershberger-testen til brug for foreløbig risikovurdering og fareklassificering.

Pubertets-assays, den intakte han-assay, og den udvidede OECD TG 407 fremskaffer viden om et stofs potens in vivo. Effekter på de mange forskellige endpoints, som er inkluderet i disse tests, kan anses for at være skadelige og/eller repræsentere en effekt på en mekanisme, som er relevant for mennesker. Derfor kan disse tests anvendes til at fremskaffe NO(A)ELs/LO(A)ELs til brug for human risikovurdering. Anvendelse af en større sikkerhedsmargin end den, der anvendes baseret på data fra OECD to-

generationsstudiet, bør overvejes, fordi toksiske effekter på udviklingen i adskillige tilfælde har vist sig at forekomme ved lavere dosisniveauer end dem, der medfører effekter i disse tests.

Reproduktions- og udviklingstoksicitetsundersøgelser tilvejebringer data om skadelige effekter, og kan bruges til risikovurdering og fareklassificering, da resultaterne indikerer en mulighed for effekter i mennesker. En række mulige forbedringer af de eksisterende guidelines synes relevante i forhold til at kunne påvise hormonforstyrrende effekter. Fravær af effekter i de eksisterende reproduktionstoksicitetsundersøgelser, som for eksempel OECD's to-generationsstudie, kan derfor ikke på nuværende tidspunkt udelukke muligheden af hormonforstyrrende effekter. De effekter, der ses i reproduktionstoksicitetsundersøgelser, kan skyldes andre mekanismer end en hormonforstyrrende, men mønstret af effekter kan indikere, at hormoneffekter er involveret. For eksempel kan et mønster med nedsat anogenital afstand, mangelfuld tilbagedannelse af brystvorter og effekter på reproduktionsorganerne hos det hanlige afkom indikere, at anti-androgene effekter er involveret, mens tidlig kønsmodning hos hunner i fravær af effekter på kropsvægten indikerer, at østrogene effekter er involveret. I sådanne tilfælde kan resultaterne bruges til optagelse af stoffet på EU's liste over mulige hormonforstyrrende stoffer.

Forslag

På baggrund af overvejelser vedrørende følsomheden i forhold til hormonforstyrrende effekter, foreslås forbedringer af de eksisterende langtids OECD-guidelines for reproduktions- og udviklingstoksicitet. Behovet for disse forbedringer skal overvejes i relation til den anvendte teststrategi. De foreslåede forbedringer er:

- TG 407, 28-dages oralt toksicitetsstudie i gnaver med gentagen dosering (enhanced): Medtag stadiebestemmelse af sædceller i den histopatologiske undersøgelse af testes som kompensation for ikke at dosere gennem en fuld sædcellecyklus.
- TG 414, prænatal udviklingstoksicitet: Overvej at indkludere målinger af testosteron niveauer i testes hos fostre på dag 21 i fostertilstanden.
- TG 415, et-generationsstudie: Opdater ved at inkludere de samme endpoints som i to-generationsstudiet, for eksempel sædanalyser, østruscyklus og histopatologi hos forældredyr. Desuden, overvej at udvide eksponeringsperioden til dag 90 efter fødslen i stedet for dag 21, og inkluder vurdering af anogenital afstand, tilbagedannelse af brystvorter, sædkvalitet og østruscyklus, og histopatologiske undersøgelser af reproduktionsorganerne hos afkommet.
- TG 416, to-generationsstudie: Inkluder vurdering af anogenital afstand og tilbagedannelse af brystvorter i F1 og F2 samt undersøgelser af misdannelser af reproduktionsorganerne hos mere end et afkom pr. køn pr. kuld.
- TG 426, Udviklingsneurotoksicitet: Vurder effekter på adfærdsforskelle mellem kønnene og inkluder vurdering af f.eks. parringsadfærd.

I forhold til fareklassificering er det blevet foreslået af WWF, at hormonforstyrrende stoffer skulle placeres i hormonforstyrrende underkategorier afhængigt af graden af tilgængeligt bevis for hormonforstyrrende effekt. Vi finder, at udvidet brug af de

tilgængelige data om stoffers hormonforstyrrende effekter kan være mere realistisk og relevant. I forhold til fareklassificering for reproduktionstoksicitet foreslår vi derfor:

- Viden *fra in vitro* testning eller *in vivo* screening kan anvendes som "anden relevant information", der viser, at det kemiske stof virker med en mekanisme, som er relevant for mennesker. Denne information kan støtte en "opgradering" fra reproduktionstoksisk, kategori 3 til kategori 2 i de tilfælde, hvor det diskuteres om de reproduktions- og udviklingstoksiske effekter skal anses for at være skadelige.
- Viden fra *in vivo* screening modeller som Hershberger og Uterus-testen kan anvendes direkte til fareklassificering, fordi positive testresultater indikerer reproduktions- og udviklingstoksicitet i undersøgelser som OECD's to-generationsstudie ved dosisniveauer af tilsvarende eller lavere størrelsesorden. Som følge heraf synes det påkrævet, at sådanne kemiske stoffer klassificeres for reproduktionstoksicitet mindst i kategori 3.

En opgradering fra kategori 3 til kategori 2 betyder, at det kemiske stof vil blive omfattet af autorisationsordningen. Imidlertid vil kemiske stoffer, der er indplaceret i kategori 3 baseret på in vivo viden om hormonforstyrrende aktivitet og en stærk mistanke om mulig udviklingstoksicitet, ikke automatisk blive omfattet af krav om autorisation. På baggrund af den stærke mistanke om udviklingstoksiske effekter mener vi, at autorisation af sådanne kemiske stoffer bør overvejes.

1 Introduction

1.1 Background and definition

There is growing concern on possible harmful consequences of exposure to chemicals that are capable of modulating or disrupting the endocrine system. This concern for endocrine disrupting chemicals (EDCs) is considering both wildlife and humans. Concerns regarding exposure to EDCs are primarily due to adverse effects observed in certain wildlife, fish, and ecosystems; increased incidence of certain endocrine-related human diseases; and endocrine disruption in laboratory animals exposed to certain environmental chemicals. This has led to a series of stakeholders, including the European Commission, to consider the topic of endocrine disruption as of sufficient concern to justify action.

The Nordic countries are also concerned that EDCs may be harmful to humans and the environment and therefore find it important that EDCs will be covered properly by the chemicals legislation in force. Therefore, the Nordic co-ordination group for test method development (Nord-Utte) has initiated a project with the aim to investigate how the OECD CF can be used as a basis for regulatory initiatives towards endocrine disrupters.

Endocrine disrupters are defined by OECD (2002a) in a generic sense as follows:

"An endocrine disrupter is an exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub) populations."

"A potential endocrine disrupter is an exogenous substance or mixture that possesses properties that might be expected to lead to endocrine disruption in an intact organism, or its progeny, or (sub)populations."

It is implicit in the definitions that a chemical can only be definitively considered an endocrine disrupter on the basis of an *in vivo* model, where a functional endocrine system is present and where a full interplay between normal physiological and biochemical processes can occur. However, it is accepted that it is possible to identify potential endocrine disrupters using other types of models (OECD, 2002a).

It should be implied that all glands, tissues, receptors, transport proteins and enzymes involved in development and normal body function may be targets for toxicity of EDCs (Harvey and Johnson, 2002).

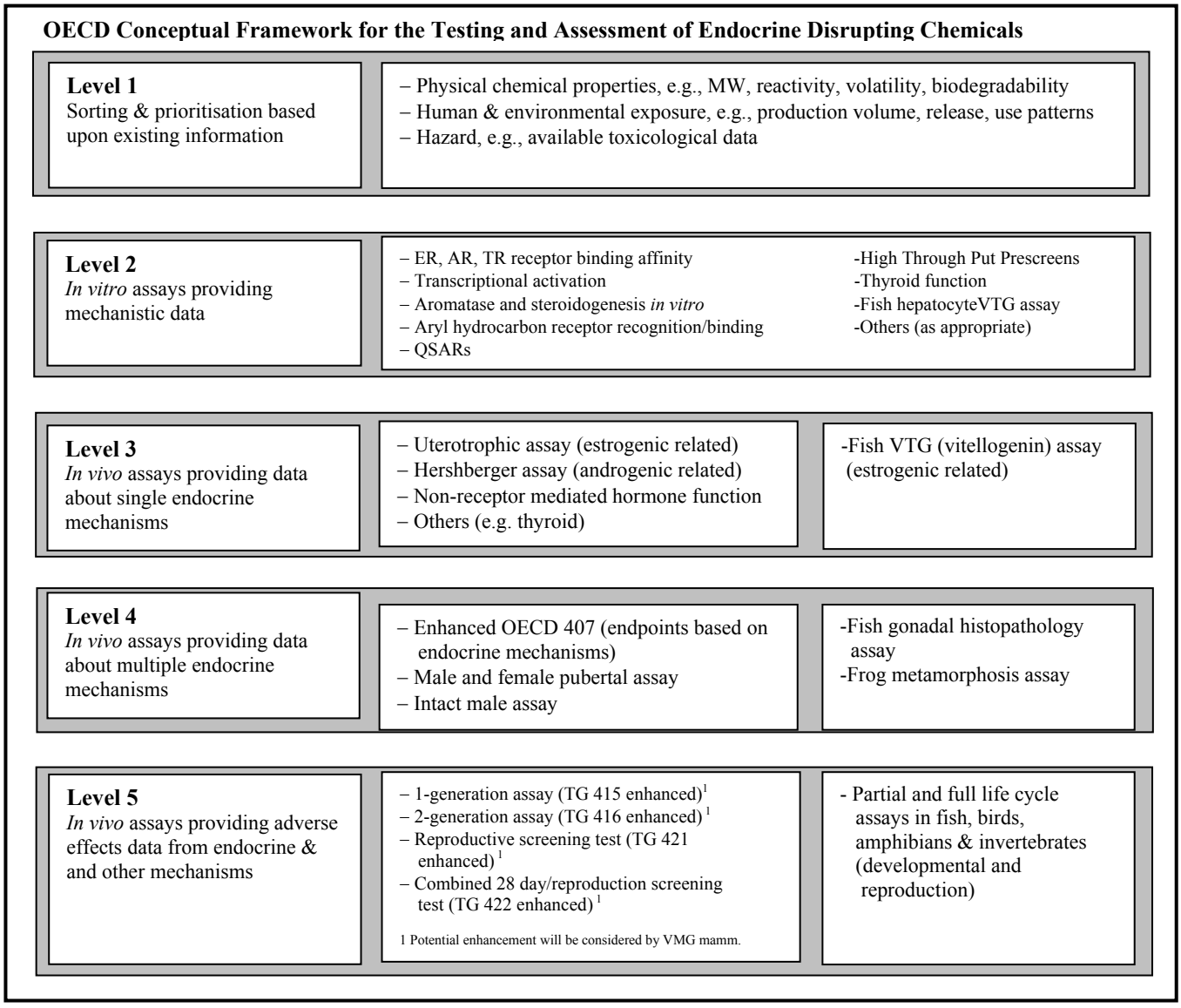
1.2 OECD for the Testing and Assessment of Endocrine Disrupting Chemicals

The OECD Task Force on Endocrine Disrupters Testing and Assessment (EDTA) of the Test Guideline Programme has during the recent years developed a Framework for Testing and Assessment of Endocrine Disrupting Chemicals. At the 6th EDTA meeting in Tokyo June 2002, the OECD Conceptual Framework for the Testing and Assessment

of Endocrine Disrupting Chemicals has been reconsidered and substantially revised. The Task Force reached consensus on a framework of 4 compartments or "levels", each characterized by the type of information it generates and 1 additional level that is used to collect all tools for sorting and prioritisation. The Task Force agreed that the framework should be kept as simple as possible, but in keeping it simple there was a need identified for a number of notes to the framework that should be considered as an integral part of it. The agreed Conceptual Framework, comprising 5 levels, together with 6 notes, is shown in Figure 1.

The Conceptual Framework is not a testing scheme, but rather a toolbox where the various tests can contribute with information for the detection of the hazards of endocrine disruption. The toolbox is organised into a number of compartments each corresponding to a different level of biological complexity (for both the toxicological and ecotoxicological area). Although the toolbox comprises a number of tests, it was recognized that it is not necessary to have data from all of them in order to evaluate a chemical substance in relation to ED properties.

Figure 1. Conceptual Framework (OECD, 2002b)



Notes to the conceptual framework

Note 1: Entering at all levels and exiting at all levels is possible and depends upon the nature of existing information needs for hazard and risk assessment purposes

Note 2: In level 5, ecotoxicology should include endpoints that indicate mechanisms of adverse effects, and potential population damage

Note 3: When a multimodal model covers several of the single endpoint assays, that model would replace the use of those single endpoint assays

Note 4: The assessment of each chemical should be based on a case by case basis, taking into account all available information, bearing in mind the function of the framework levels.

Note 5: The framework should not be considered as all inclusive at the present time. At levels 3, 4 and 5 it includes assays that are either available or for which validation is under way. With respect to the latter, these are provisionally included. Once developed and validated, they will be formally added to the framework.

Note 6: Level 5 should not be considered as including definitive tests only. Tests included at that level are considered to contribute to general hazard and risk assessment.

The idea behind the establishment of the different levels is that it may be possible to use information at each level, make an assessment and exit at the same level for risk management in relation to existing data. This means that different legislative instruments can be used with regard to EDCs depending on existing data and the level the assessment has been based on. When new data become available, revised hazard and risk assessment and subsequently risk management actions can be considered. Using this approach, regulatory authorities will not have to await all test results from a tiered test program, but may be able to consider immediate action based on the available data.

1.3 EU Chemicals Policy

In 2001, the European Commission (EC) issued a White Paper entitled "Strategies for a Future Chemicals Policy", which was subsequently endorsed by the Member States of the European Union (EU) (CEC 2001).

Existing chemicals (i.e. chemicals that were already on the market before new EU legislation on chemicals came into force September 1981) have been evaluated differently from that adopted for new chemicals (i.e. chemicals marketed after September 1981). Since the testing requirements for existing chemicals is less rigorous than for new chemicals, there is concern that a substantial number of the existing chemicals that are currently marketed, may have been inadequately tested and could therefore be harmful. To address this problem, the EC White Paper proposed the establishment of a new system called REACH (Registration, Evaluation and Authorisation of Chemicals).

The European Commission's proposed new chemicals legislation states that chemicals meeting certain criteria for very high concern should be brought within an authorisation scheme. According to the EU white paper, endocrine disrupting chemicals should be classified in accordance with the existing chemicals legislation. In June 2001, the Environment Council concluded that EDCs should be covered by the authorisation procedure in the new EU chemicals regulation, REACH, when scientifically valid test methods and criteria have been established.

The EU Commission launched the proposal for REACH in October 2003 (EU 2003). In the Commission proposal endocrine disrupters have been included in the authorisation procedure based on a case-by-case assessment. In case a member state evaluates that a substance should be covered by the authorisation procedure due to endocrine disrupting properties, the member state shall elaborate a dossier and send to the Agency for further action.

Several chemical substances with endocrine disrupting properties are carcinogenic, toxic to reproduction or the environment or have other properties that have lead to classification according to the existing regulation. At present, it is uncertain whether or to what extent these effects are caused by the ED properties. In addition and more importantly, it is uncertain to what extent the existing classification criteria will lead to classification due to some of the effects that are of relevance for EDCs, e.g. disturbances of the development of reproductive organs and functions, thyroid effects, and effects on sexual dimorphic behaviour.

As endocrine disrupters in the future should be handled based on a case-by-case assessment and as long as scientifically valid test methods and criteria have not been developed, there is a need for guidance in doing this assessment – how to interpret test

results and identify a substance as an endocrine disrupter. Furthermore, development of test methods and criteria may take some time and meanwhile there is a need for considering regulatory initiatives in relation to EDCs

1.4 Globally Harmonised System of classification and labelling of chemicals (GHS)

A globally harmonised system of classification and labelling of chemicals have recently been agreed upon. This means that in the coming years, the EU classification system for chemicals will be changed so the EU legislation will be in line with the GHS-system. Even though there are a many similarities between the two systems, introduction of the GHS-system will probably cause some changes. In this report, proposals for regulatory initiatives with regard to hazard classification are related to the existing EU-classification. The existing criteria for classification and labelling of substances toxic to reproduction are listed in table 12 whereas the GHS-criteria is listed in table 13. However, the proposals can easily be transferred to the GHS-system – in general category 1, 2 and 3 in the EU criteria can be transferred to category 1A, 1B and 2 in the GHS-system, respectively.

1.5 Aim of the project

The main purpose of the project is to consider the OECD Conceptual Framework for Testing and Assessment of Endocrine Disrupting Chemicals in relation to regulatory initiatives for EDCs. The report will also serve as a valuable guidance in how to interpret test results and how to identify a substance as an endocrine disrupter. Hopefully, this may serve as a valuable support for the future case-by-case assessment of endocrine disrupters in EU.

In general, the report is expected to serve as a valuable and important basis for the Nordic contribution to the discussions in EU about how to interpret and use test results that indicate endocrine disruption for regulatory purposes, and furthermore how to integrate endocrine disrupters in the new EU chemicals regulation.

The aims of the project are to provide an assessment of the tests in the OECD Conceptual framework, including specification of the endpoint for the test and reliability and relevance for effects in humans as well as proposals for legislative instruments to be used when data are available at a specific level.

Thus, the objectives are to:

- consider endpoints included, reliability of results and relevance for humans
- consider gaps/deficiencies in existing OECD reproductive toxicity test guidelines
- identify possible enhancements of existing OECD guidelines
- propose regulatory actions for EDCs in relation to human health

The project is focused on human health effects, especially effects of oestrogenic and androgenic agonistic or antagonistic activity with regard to effects on human reproduction. Other forms of endocrine disruption involving other hormonal systems will be included to a limited extent for thyroid effects, while the adrenal system fall

outside the scope of this document. In the future, it may be considered to perform similar work with regard to other health effects, e.g. cancer, and environmental effects.

The first part of the project includes an assessment of each type of test at each level and considers the endpoints with focus on their reliability and relevance for effects in humans. The second part considers the possibility for evaluating a chemical as having endocrine disrupting properties based on test data on a specific level, and proposals for regulatory actions in practise.

Examples of legislative instruments that can be considered are: prioritisation for the EU-list of potential endocrine disrupting chemicals, prioritisation for further testing, risk reduction based on concern identified via risk assessment, hazard classification with a Risk-sentence, or inclusion of the substance in the authorisation procedure of REACH.

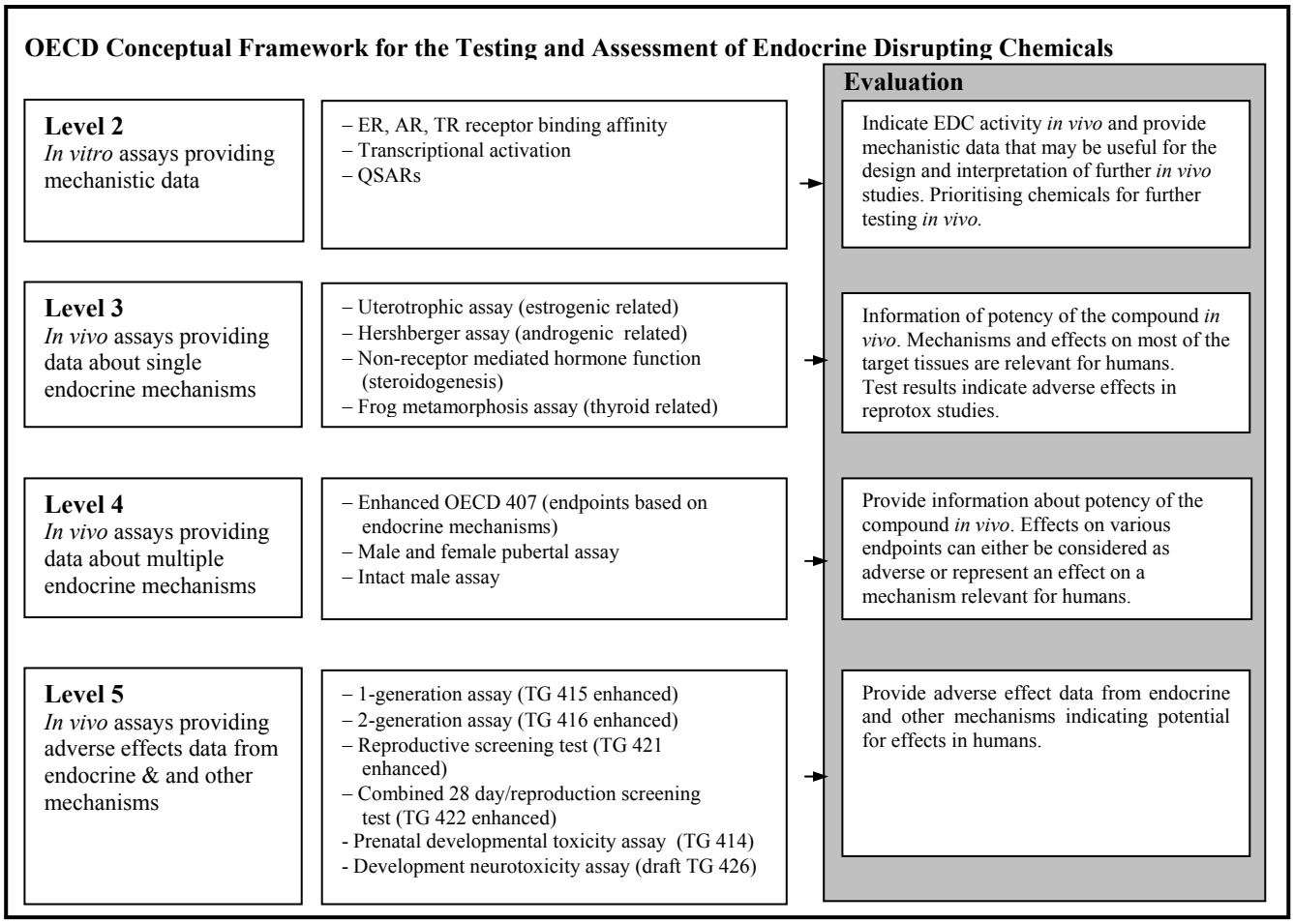
2 Conceptual Framework – with specific focus on level 2-5 with regard to toxicity testing

2.1 Introduction

In this chapter, the assays in the OECD Conceptual Framework will be considered with special emphasis on levels 2-5. A modified version of the original Conceptual Framework is represented in figure 2, which includes the assays considered in this section.

A specification and evaluation of the endpoints for the test at each level regarding reliability and relevance for predicting effects in humans will be given together with a discussion for each assay with specific focus on advantages, limitations and deficiencies. Furthermore an overall conclusion can be found at each level.

Figure 2. A modified version of the Conceptual Framework including the assays evaluated and the evaluation at each level (OECD, 2002b).



2.2 Level 2, *in vitro* assays providing mechanistic data

2.2.1 Introduction

In vitro tests are used to test the ability of a given chemical to bind to or mediate response from the androgen receptor (AR) or the oestrogen receptors (ER) and act like an agonist or an antagonist. The following will focus on the purpose with in vitro testing and the advantages and limitations. These considerations are mainly based on a review by Earl Gray et al. (1997) and the OECD Detailed Review Paper (2002a).

The purpose of *in vitro* assays is to detect endocrine activity e.g. the ability of a substance to bind and active or block the ERs and AR, but not to determine dose-response relationship. It is acknowledged that *in vivo* testing is needed for identifying endocrine disrupting chemicals more definitively, however, these tests may be long-term and more expensive studies. For these reasons, it is under consideration whether in vitro tests could be utilised to screen a large number of environmental chemicals for EDC activity.

Generally, the advantages of *in vitro* methods include:

- Identification of mechanism of action
- Sensitivity to low concentrations increases detectability

- Specificity of response
- Low cost and short time
- Small amount of material required
- Testing can be automated
- High throughput assays can be developed
- Reduced use of experimental animals

The advantages and limitations of some *in vitro* assays for EDC activity are shown in tables 1 and 2, while advantages and limitations of QSAR models are shown in table 3.

2.2.2 Discussion

Although *in vitro* methods could be used for screening for EDC activity, the need for in vivo testing is important to consider. A vast number of in vitro tests would be required to screen for all EDC activities due to the multiple mechanism by which EDCs may act (i.e. altering hormone synthesis, transport, receptors, metabolism). However, many endpoints in i*n vivo* experiments are required to cover all effects as well.

For *in vitro* assays, the metabolic capacity of cultured cells and the solubility of chemicals in media are important considerations. The inability of many, but not all, *in vitro* systems to metabolically activate toxicants is a major limitation of these methods, because this may give rise to false negative results.

In cases where it is known that certain classes of chemicals do not require metabolic activation or deactivation, or the metabolites are known and tested, some *in vitro* tests may offer advantages over *in vivo* screening, because the *in vitro* tests generally are faster.

At present, it is unclear to what extent *in vitro* data would be useful for risk assessment, because in vitro potency does not always correlate with *in vivo* toxicity due to mechanistic and kinetic factors. When comparing *in vitro* MCF-7 data with results from *in vivo* test systems, Mayr et al (1992) found a similar order of relative potencies comparing exoprotein induction in MCF-7 cells with the uterotrophic assay in mice when testing mycoestogens and phytooestrogens (OECD, 2002a). However, the *in vitro* assay tended to indicate higher activities, the lower activities *in vivo* most likely being due to compound metabolism.

Table 1. *In vitro* assays for oestrogenic activity

Assay	Design	Advantages	Limitations
In vitro ER binding assays	Cytosolic or nuclear extracts containing ERs are incubated with radio-labelled E2 in the presence or absence of increasing concentrations of test chemicals. Specific binding is measured.	Fairly inexpensive. Measure ER in cell-free extracts from tissue from exposed and controls. Measure competitive binding with E2 to the ER. Sensitive. Short duration, can be standardised between laboratories.	Does not distinguish between ER agonist and antagonist. May give false negative results if metabolic activation is required prior to binding to ER. Not always comparable between labs, but should be possible to standardise. Uses rat ER.
MCF-7 cell assay, ER binding	Uses cell lysate to measure affinity under cell-free conditions. Competition of chemical with radiolabelled E2 for specific binding to the ER.	Bioavailability and metabolic activity of a chemical can be evaluated. Uses human ER. Metabolically activate some prooestrogens.	Does not distinguish between ER agonist and antagonist.
MCF-7 cell assays, cell proliferation assays (E-SCREEN assay)	Oestrogen-specific cell growth.	Sensitive. Reproducible. Can distinguish agonist and antagonists.	Requires optimisation of various laboratory and culture conditions and may be difficult to standardise for large-scale testing. Proliferative response differs between substrains of MCF-7 cells. Takes longer time than other *in vitro* assays (6 days). False positives (general cell mitogens). False negatives (cytotoxic, general growth inhibitors).
Transiently transfected ER-mediated transcription assays in MCF-7 cells	Utilise the hER for transcriptional regulation of a reporter gene. The most sensitive assays uses a luciferase reporter gene.	Identifies oestrogenic chemicals. Simple assay protocol allows screening of large numbers of chemicals. Reliable and reproducible method. A sensitive evaluation of a compounds ability to induce oestrogen-regulated transcription. A chimeric hER is only activated by ER ligands. Can rapidly distinguish between ER agonists and antagonists. High sensitivity. Specific.	Only some oestrogenic chemicals induce transcription. Transfections have to be performed to each experiment.
Stably transfected ER transcription assays in MCF-7 cells	Assay based on MCF-7 cell derivative containing an ER-controlled segment of a promoter gene that regulates the expression of e.g. luciferase activity Examples are the MVLN and the ER-CALUX assays and MCF-7 cells transfected with ERE-LUC.	Easy to use because cells are permanently transfected. Short term assay. Oestrogen-regulated transcription measurable with high sensitivity. Standardised assay. Detection of oestrogen agonists and antagonists. Specific.	Maybe difficult to establish a stable transfected cell line. Cells may lose activity over time.
ER-mediated transcription assays in yeast cells	Mammalian steroid receptors introduced into the yeast strain functions as steroid-dependent transcriptional activators Cells are stably transfected with hER and a reporter gene.	Easy to use. Short-term duration. Ability to quantify without using radioactive materials.	Does not distinguish between ER agonist and antagonist. Phylogenetic differences in metabolism may exist. Yeast cells have cell wall, and porosity and active transport mechanisms may vary from mammalian cells. Low resolving power for low potency chemicals.

(Earl Gray et al.,1997; Legler et al., 2003; OECD 2002a, Vinggaard et al., 1999a, Fang et al., 2000).

Table 2. *In vitro* assays for androgenic and antiandrogenic activity

Assay	Design	Advantages	Limitations
In vitro AR cell-free binding assays	Ability of chemicals to compete with endogenous ligand for binding to AR.	Fairly inexpensive. Sensitive. Short duration. Can be standardised between laboratories.	Does not distinguish between AR agonist and antagonist. May give false negative results if metabolic activation is required.
Whole cell AR binding assays	Ability of chemicals to compete with endogenous ligand for binding to AR.	Metabolically activates some proantiandrogens. Reproducible. Easy to perform.	Expensive and time consuming.
Reporter gene assay transient transfected with AR	Based on transiently transfected Chinese hamster ovary cells. Determines AR agonists and antagonists by competitive binding of steroid and chemical to hAR. Chemical action is measured by inhibition or activation of luciferase activity.	Uses hAR and display some metabolic activity. Distinguishes between agonists and antagonists. Very high sensitivity. Some metabolic activity. Detects and quantifies AR mediated androgenic and antiandrogenic effects of chemicals with reasonable accuracy.	The assay can be complicated to perform.
Reporter gene assay stably transfected with AR	Based on stably transfected human prostate carcinoma cells or human mammary carcinoma cells.	Distinguishes between agonists and antagonists. High sensitivity. Detects and quantifies AR mediated androgenic and antiandrogenic effects of chemicals with reasonable accuracy. Some metabolic activity.	Cells may lose activity over time.
Cell proliferative assay (A-SCREEN)	Based on mammary carcinoma cells stable transfected with hAR (MCF-7 AR1). Anti-androgens inhibit cell proliferation.	Distinguishes between agonists and antagonists. High sensitivity. Detects and quantifies AR mediated androgenic and antiandrogenic effects of chemicals with reasonable accuracy. Some metabolic activity.	Less dynamic range than other reporter gene assays which means this assay shows less induction compared to the other tests.
AR DNA binding assay	Chemicals that bind AR and inhibit DNA binding Reduces the intensity of the bands in the band-shift assay in a dose-response manner.	Provides additional information about the mechanism of action for antagonists.	Provides a large number of data-more than needed for regulational purposes. Assay can be complicated to perform.
Yeast-based AR assay	Mammalian AR introduced into the yeast strain functions as steroid-dependent transcriptional activators.	Ease to use. Short-term duration. Ability to quantify without using radioactive materials.	Does not distinguish between AR agonist and antagonist. Phylogenetic differences in metabolism may exist. Yeast cells have cell wall, and porosity and active transport mechanisms may vary from mammalian cells.

(Earl Gray et al.,1997; OECD 2002a; Szelei et al., 1997; Térouanne et al., 2000; Vinggaard et al., 1999b and Wilson et al., 2002)

Table 3. Other assays

Assay	Design	Advantages	Limitations
Quantitative structure-activity relationship (QSAR) models	Based on regression analysis, neural net and classification approaches and on clustering of chemicals by functional relationship and biological properties.	Reduces costs. Reduces initial use of animals. May speed up decision making.	Needs to be assessed for reliability and uncertainties in predictions of reproductive toxicity.

The cell free tests for oestrogenic or androgenic activity are able to show whether a given chemical can bind to a particular receptor or not, with the inference that a high binding affinity would potentially results in marked biological activity of some type. However, these test systems do not distinguish between agonist and antagonist activity, like the reporter gene assays (Vinggaard et al., 1999).

An approach that indirectly involves biological material and data from such are Structure-Activity Relationships or Quantitative Structure-Activity Relationship – together abbreviated (Q)SAR. The underlying hypothesis for the models is that chemical substances with similar structures will have similar properties. A (Q)SAR is a relation between structure properties of chemical substances and another property. This can be a physical-chemical property or a biological activity, including the ability to cause toxic effects.

QSAR models are in the EU White Paper on a new chemicals policy believed to be important in future chemical management, including priority setting, classification and risk assessment (CEC, 2001). In 2002, the OECD started up a work program to promote the regulatory acceptance and use of (Q)SARs to fill data gaps and thereby reduce animal use and costs. In the EU, there is general agreement that objectives envisioned under the proposal for a new chemicals legislation, REACH, can only be achieved with the help of (Q)SAR techniques.

QSAR results have been used as the basis for an advisory list for self-classification of dangerous substances prepared by the Danish Environmental Protection Agency in 2001 (MST 2001) It was stated by the Danish EPA that the (Q)SAR models used for identification of chemicals with dangerous properties e.g. mutagenicity or carcinogenicity are now so reliable that a substance can be predicted with an accuracy of 70-85%.

Models for reproductive toxicity represent, however, a special challenge due to the diversity of causative factors within the multitude of test systems, including long term in vivo assays, where many different endpoints are assessed. There is at present no definitive high quality data set available for effects found in long term assays concerning reproductive toxicity with which to develop a reliable global model1. This is not to say that there are no models, which may be useful in predicting reproductive effects in certain cases. A model like e.g. the teratogenicity model from Multicase Inc., which is derived on human data, may occasionally be of some use in predicting human teratogenicity.

[1] A global model is a model which is able to make predictions of a large fraction of ”the chemicals universe”

For some types of chemicals sufficient test data do exist to make useful models that may predict certain types of toxicity, which may also relate to reproductive toxicity, e.g. estimates of Rodent Dominant Lethal or Drosophila Melanogaster Sex-Linked Recessive Lethal effects[2] (pers. com., Danish EPA). In cases where specific mechanisms have been identified it may also be possible to develop and make use of predictive models e.g. with regard to hormonal effects such as anti-androgenicity.

2.2.2 Conclusions at level 2

Positive *in vitro* test results indicate potential ED activity and a potential for ED effects *in vivo*. *In vitro* data can provide valuable mechanistic data that is useful for the design of further *in vivo* studies. The *in vitro* tests are relevant for effects in humans because many of these tests are based on human hormone receptors. Chemicals that bind to these receptors are therefore likely to cause effects in *in vivo* studies and on reproductive function in humans.

Negative *in vitro* test results cannot be used to exclude potential EDC activity because of limitations such as inability or unknown capacity to metabolically activate toxicants and because EDC activity can occur through mechanism other than those tested in the *in vitro* test system.

QSAR models for ED activity and reproductive toxicity effects are under development, but at present the use for priority setting and risk assessment is undecided

2.3 Level 3, Short terms in vivo assays providing data about single endocrine mechanisms and effects

This level includes the Uterotrophic and the Hershberger assay, as well as assays for non-receptor mediated hormone function and other assays providing data about single endocrine endpoints.

The endpoints include alterations in the accessory reproductive organs, testes, ovaries, uterus, adrenals, thyroid gland, hypothalamus and the pituitary; all target organs for EDCs. These organs are all endocrine related inter playing via feedback mechanisms and interacting with development and function of a diverse range of target organs and tissues.

To be able to address a chemical as an EDC ideally all these organs and tissues should be evaluated in testing of chemicals. However, this would involve a major task that in most cases would not be feasible.

The endpoints reveal interference by a chemical with one pathway of action. When the tests are designed to address a particular pathway, further testing is needed to address all the endocrine pathways that could be affected by an EDC.

The Uterotrophic and the Hershberger assay are at present close to being validated within OECD and the guideline preparation is in progress. The other assays at level 3 are not validated yet, however, results from these tests may provide information on the action of a chemical on the endocrine system.

[2] these two tests are normally categorised as germ cell mutagenicity tests, but they may also give some indication on embryo toxic potential

2.3.1 **Uterotrophic** assay

The Uterotrophic assay is an *in vivo* test method for detection of chemicals that have the potential to act like and interfere with the endogenous female sex hormone. The purpose of this short term *in vivo* screen is to identify suspected oestrogen (ant)agonist. Design and endpoints are listed in table 4.

The principle underlying the Uterotrophic assay is that the uterus growth phase is under control of oestrogens in the natural oestrous cycle. Oestrogens are necessary to stimulate and maintain growth of the uterus. When endogenous source of oestrogen is not available, either because of immaturity or because the animal is ovariectomized, the animal will require an exogenous source of oestrogens to restore and maintain growth of the uterus. Chemicals that act as agonists, can be identified if they cause an increase in uterus weight. If chemicals decrease uterus weight when co-administrated with a potent reference oestrogen, they will be identified as antagonists. The design and endpoints are listed in table 4.

Uterus weight
The uterus is an oestrogen sensitive tissue. Several studies have shown that hormones and endocrine disrupting chemicals can have an effect on uterus weight in the Uterotrophic assay. The assay has been used since 1935 and a protocol has recently been standardised within the OECD TGP (Owens and Ashby 2002). The oestrous cycle in the rat is 4-5 days, so the 3 days administration of test compound used in the OECD protocol is similar to the response time to endogenous oestrogen surge that stimulate the uterine tissue in the intact animal. The OECDvalidation study demonstrates that the uterus weight is a relevant endpoint for identifying both potent and weak oestrogen agonists and potent antagonists, when using either intact immature or ovariectomized adult female rats. This is despite of different rat strains and routes of administration (the most sensitive route of administration might vary depending on the chemical being tested) (OECD, 2003a).

2.3.1.1 Discussion

Specificity of *in vivo* assays: the target tissue in the Uterotrophic can be affected by other mechanism than endocrine disruptions. The uterine weight in ovariectomized rats can be affected in an oestrogen like manner by high doses of androgens and growth factors. If intact females are used in the Uterotrophic assay it is possible that substances that act via the hypothalamus-pituitary-gonadal axis can affect the uterine weight (Gray et al. 1997)

The effects of oestrogen on uterus are mediated by oestrogen receptors (ERs) ER-alpha and ER-beta. Ligand binding studies indicate that estradiol has comparable affinity for the two ERs (Kuiper et al. 1997). *In situ* hybridisation histochemistry studies in rat uterus reveal that ER-alpha is highly expressed in rat endometrium compared to ER-beta (Shughrue et al. 1998). Furthermore studies with ER-alpha knocked-out mouse show that uterus in these animals are not responsive to alterations in oestrogen levels (Couse et al. 1997). These studies indicate that ER-alpha is the predominant ER in uterus and that ER-beta does not play a critical role in the oestrogen regulation of uterus.

A number of molecular and biochemical's markers, histological change and mitotic events have been suggested as additional endpoints in the Uterotrophic assay. Several studies have included additional endpoints in the Uterotrophic assay to compare these

responses with the change in uterine weight. It has been considered that additional endpoints in the Uterotrophic assay will enhance the sensitivity of the test and thereby improve the usefulness in identifying potential oestrogen substances. For example, measurements of the cell height of the luminal epithelial in uterus were studied by Markey et al. (2001), and Tinwell et al. (2000). In the first experiment an increase in the height of the luminal epithelium was shown to be a more sensitive endpoint than an increase in the uterine weight, whereas the second experiment found statistical significant increase in uterine weight at dose level without a parallel increase in epithelia cell height. Both studies used bisphenol A (BPA). A review by Owens and Ashby (2002) indicates that a comparison of Uterotrophic studies which include additional measurements of oestrogenic activity show some inconsistencies. Where the same additional endpoint may respond prior to the uterine weight increase at a lower dose in one study, it responded at the same level in another study, and did not respond at all in a third study. Whether these additional endpoints are more or less sensitive than the uterine weight change is thus not evident at present.

Laws et al. (2000) have compared the Uterotrophic assay with other test models for detection of oestrogenic activity which include: the age of vaginal opening, the induction of cornified vaginal epithelia cells in ovariectomized adult rats and oestrous cyclicity in intact adult rats. The substances tested were: estradiol, methoxychlor, nonylphenol (NP), octylphenol (OP), and BPA. The results from this study indicate that the Uterotrophic assay was the most sensitive and consistent method and that the vaginal cornification assay was the most insensitive in detecting oestrogen effect.

A review by Owens and Ashby (2002) indicate that there is a general correspondence between results in the Uterotrophic assay and the outcomes in testing for adverse effects in reproductive and developmental toxicity assays, when the same route of exposure is used with NP, OP, BPA, Methoxychlor and genistein as test chemical. Several compounds positive in the Uterotrophic assay have shown effects consistent with oestrogen mode of action in reproductive and developmental toxicity assays, when similar or higher doses were used. A few compounds which were weakly positive at only high dose levels in the Uterotrophic assay did not show apparent oestrogen-related effects in reproductive and developmental toxicity assays. Test substances that were negative in the Uterotrophic assay have not elicited any oestrogen-related effects in reproductive and developmental toxicity assays at similar dose levels. Generally, a limited occurrence of false positives in a screening assay is to be preferred rather than the occurrence of false negatives.

It has been argued that weight changes of the hormone responsive tissues in the Uterotrophic assay shows a mechanistic effect and is therefore only a tool for the prioritisation of chemicals for further testing. The present use of this assay is to demonstrate the agonistic or antagonistic characteristic of chemicals *in vivo* in preliminary screening assays. A positive response in the Uterotrophic assay has been claimed not in itself to be an adverse effect but rather to be indicative of other and adverse properties of the chemical (Newbold et al. 2001). A positive response in this assay suggests the need for the substance to advance to reproductive and developmental testing for adverse effects. A negative uterotrophic response, in a thorough doses-responds study, shows that the test compound is not an ER ligand.

In conclusion, the Uterotrophic assay appears to be reliable in identifying substances with an oestrogenic mode of action. Clear evidence supports that an increase in uterine

weight normally is caused by an oestrogen mode of action. The sexually immature assay and the ovariectomized mature assay have shown equivalent results in multi-laboratory comparison study within OECD TGP. General correspondence between results in the Uterotrophic assay and the outcomes in testing for adverse effects in reproductive and developmental toxicity assays exist. The mechanisms involved in these assays are highly relevant for humans. Likewise, the effects on target tissue are relevant for humans.

Table 4. Overview of the Uterotrophic and Hershberger assays

Test	Design	Endpoints	Advantages/Limitations	Guideline(s)
Uterotrophic Assay	3 day s.c. or p.o. administration to either intact immature or adult ovariectomized female rats to detect (anti)oestrogens (n=6/group)	Mandatory endpoints: body weight and weight of oestrogen responsive tissue: Uterus wet weight and blotted weight.	+Used since 1935. +Simple, robust and reproducible. +Test result correspondence to reprotox. tests. -Detects ER ligands only . -Inhibition of steroidogenesis not detected. -Uses live animal.	In preparation: The first and second phase of the validation of the Uterotrophic assay within OECD is completed. The guideline preparation is in progress within OECD.
Hershberger Assay	10 day p.o. administration to immature castrated male rats to detected (anti)androgens (n=6/group)	Mandatory endpoints: body weight and weight of: ventral prostate, seminal vesicles plus coagulating glands, levator ani/bulbocavernosus muscle, Cowper's glands, glans penis. Optional endpoints: weight of liver, kidneys and adrenal glans and measurements of serum testosterone and LH hormone levels.	+Used since 1940. +Simple, robust and reproducible. +More sensitive than the pubertal assay to AR ligands. -Uses live animal . -Detects AR ligands only. -Inhibition of steroidogenesis not detected. -Surgical castration.	In preparation: The first and second phase of the validation of the Hershberger assay is completed. The third phase is in progress within OECD.

2.3.2 Hershberger assay

The Hershberger assay is an *in vivo* test method for detection of chemicals that have the potential to act like and interfere with the endogenous male sex hormone. The purpose of this short term *in vivo* screen is to identify suspected (anti)androgens.

The Hershberger assay is based on the principle that a number of accessory reproductive organs require androgens to stimulate and maintain growth. If the endogenous source of androgens is not available, because the animal has been castrated, the animal requires an exogenous source to initiate and restore the growth of these tissues. Chemicals that act as agonists, will cause an increase in the weight of the androgen dependent tissues, and

as antagonists if they cause a decrease, when co-administered with a potent agonist. Design and endpoints are listed in table 4.

Weight of male reproductive accessory organs

The Hershberger assay has been used for screening and testing for hormonal activity since 1940. The assay is a relevant test method to assess chemicals with (anti)androgen activity *in vivo*. The weight response in the five androgen dependent tissues (ventral prostate, seminal vesicles/coagulating glands, levator ani/bulbocavernosus muscle, Cowper's glands and glans penis) can be different depending on different anti-androgens and their mode of action (whether they are AR antagonists or 5-alpha reductase inhibitors). Prostate differs with respect to the androgen that controls its growth and differentiation compared to the other target organs in the Hershberger assay. The prostate is dependent on enzymatic activation of testosterone to dihydrotestosterone (DHT), whereas the seminal vesicles are less dependent upon this conversion. Therefore effect on 5-alpha reductase can be distinguished from AR-antagonists by determining whether the prostate is preferentially affected. The levator ani muscle/bulbocavernosus on the other hand is testosterone dependent, having little capacity to convert testosterone to the more potent androgen DHT. Weight of this muscle is also useful in identifying anabolic androgens. This is why several target tissues are preferred and included in the assay.

2.3.2.1 Discussion

The target tissues in the Hershberger assay can be affected by other mechanism than endocrine disruptions. For instance, the weight of the male reproductive accessory organs can be affected by growth hormone.

Like in the Uterotrophic assay, additional endpoints are occasionally included in the Hershberger assay. These additional endpoints include measurements of serum testosterone and LH hormone levels. Hormone measurements are at present included as optional endpoints in the OECD Hershberger protocol and further validation will provide information about reliability of these endpoints.

Expressions of androgen-responsive genes in prostate have been included in some Hershberger studies (Nellemann et al. 2001 and 2003; Vinggaard et al. 2002). These studies indicate that gene expression analysis can be suitable for quantitative evaluation of antiandrogen function. It is still too early to predict the utility of these methods but at present these analyses can be a supplement to organ weights and hormone analysis.

It has been argued that a positive response in the Hershberger assay is not of itself an adverse effect, but the positive outcome could be indicative of other and adverse properties of the chemical. Therefore, a positive response in this assay suggests the need for the substance to advance to reproductive and developmental testing for adverse effects. A negative response indicates that the test compound is neither an AR ligand nor a 5-alpha reductase inhibitor. This means that a test compound found negative in the Hershberger assay could still have endocrine disrupting properties mediated through another mechanism.

A brief comparison between results in the phase two validation of the Hershberger assay and results in reproductive and developmental toxicity studies from the published literature have been made in order to evaluate the toxicological relevance of the Hershberger assay. For the substances: trenbolone, procymidon, vinclozolin, linuron, DDE and finasterid a good correspondence between results in the Hershberger assay

and the results in reproductive and developmental toxicity assays, were found when the same route of exposure was used (OECD 2003b).

In conclusion, the Hershberger assay, even though it is still under validation within OECD, seems reliably to identify substances with an (anti)androgen mode of action. The mechanisms involved in these assays are highly relevant for humans. Likewise, the effects on most of the target tissues are relevant for humans.

2.3.3 Non-receptor mediated hormone function

The sex hormones, e.g. testosterone, oestrogen and progesterone play a major role in the foetal development of males and females, during onset of puberty and in maintenance of the reproductive function in adults. These hormones are synthesised in the gonads and adrenals through a cascade of enzymatic events called steroidogenesis that takes place in the steroid producing cells. The production of hormones in the gonads is regulated by feedback mechanisms acting via the hypothalamus and pituitary that stimulates hormone production in the gonads via LH and FSH secretion. There is a range of different enzymes involved in steroid synthesis. Aromatase is one of these enzymes that converts testosterone to oestrogen, and 5-alpha reductase that converts testosterone to dihydrotestosterone. Some of the paths in the enzymatic cascade have been targeted by designed drugs or by chemicals blocking the enzymatic processes. These chemicals are included in the category of endocrine disrupters since they disturb the balance of endocrine hormones and thereby may cause an effect on the reproductive system.

There are multiple targets for disruption at several steps in the enzymatic cascade of steroid production. Therefore a number of endpoints should be included to elucidate possible effects. The relevant endpoints include histological examination of hypothalamus/pituitary, adrenals, testes and ovaries, measurements of steroids in serum or reproductive organs and evaluation of the lipid transport proteins and enzymes in the steroidogenic cascade.

2.3.3.1. Discussion

No short-term test has yet been proposed or included in the OECD guidelines to address chemicals with direct effect on steroidogenesis. However, there are studies where chemicals have been identified as steroidogenesis disrupters, and there are pharmaceutical products that interfere with steroidogensis. Some drug examples are formestrane (a steroidal aromatase inhibitor), anastrazole and letrozole (non-steroidal aromatase inhibitors) and aminoglutathimide (a P450 inhibitor) (Harvey and Johnson, 2002). Several chemicals have been identified as EDCs that interfere directly with steroidogenesis. One of these is lindane that inhibits steroid acute regulatory protein (StAR) (Walsh and Stocco, 2000; Harvey and Everett, 2003). The plasticiser di(2-ethylhexyl)phthalate (DEHP) reduces testicular testosterone levels in testis from in utero exposed males and inhibits ex vivo testicular testosterone production (Borch et al., 2003a). These are specific ED endpoints whereas e.g. the decreased AGD in in utero DEHP-exposed males (Jarfelt et al., in manuscript) is a non-specific endpoint. These endpoints cannot directly be attributed to a non-receptor mediated effect of the chemical.

The assays described in table 5 include measurements of steroids in serum (T, E and P) and LH and FSH levels. These are considered specific endpoints. Parts of these assays can optionally include non-specific endpoints (in terms of steroidogenic interference)

like disrupted oestrous cyclicity, increased ovarian size and abnormal ovarian or testicular histology (Gray et al., 1997). These are non-specific because they could also illustrate that the chemical has receptor binding properties or interfere via different pathways/functions than on steroidogenesis. Chemicals inhibiting aromatase in the adult male rat could affect non-specific endpoints like: disrupted mating behaviour, reduced fertility, normal hormone profile, but no effect on sperm production and storage or reproductive organ weights. In the female, aromatase inhibitors could cause ovarian alterations, delayed parturition, abnormal oestrous cycle (at high doses), reduced uterine weight and more specifically a change in serum LH concentration. Some compounds that are found to have effect in rats and mice, does not affect the guinea pig or rabbits (Hirsch et al., 1987; Gray et al., 1997; EPA, 1998). Administration of a steroidogenesis-disrupting chemical to pregnant dams during last third of gestation may affect non-specific endpoints and cause for example delayed delivery, maternal death, pregnancy loss, and more specifically inhibited oestradiol and progesterone synthesis. The particular effects depend on which enzymes are inhibited (Gray et al., 1997).

Table 5. Detection of altered steroidogenesis

Test	Design	Endpoints	Advantages/Limitations	Guidelines
Endocrine challenge test (ECT) (EPA, 1998)	Short or long term exposed intact mature animals are challenged with GnRH or an LH- or FSH-like substance that stimulates a hormonal response. Serial blood samples are collected and analysed.	Androgenic chemical: testosterone level increased. Anti-androgenic: decreased testosterone level.	+ Can determine effects on thyroid hormones. +Accepted diagnostic method in humans.	Recommended *in vivo* test by EPA, 2002 (Draft detailed review paper)
Ex vivo test to detect inhibition of steroid hormone synthesis (Gray et al., 1997)	Assessment of Leydig cell function by treatment of male rats with the chemical and assessing hormone production by the testis *in vitro* after necropsy or castration. The males can be exposed before removal of testes or the testes can be exposed *in vitro* and in both cases further stimulated by LH/hCG for several hours.	Maximum T production and LH responsiveness.	+ Metabolising system included. + Several steroid substances can be added to bypass and evaluate the affected enzyme(s). + The tissue/cell incubation procedures and hormone assays are standardised and reproducible. + Evaluates testicular steroidogenesis from GD 15 to old age.	
Ex vivo steroidogenic screen assay (EPA, 2002)	Treatment of immature or adult animals to a selected dosing regime (Route and duration of exposure is not constrained). Exposure can range from *in utero* exposure to life long exposure or short-term exposure. Testes or ovaries are dissected. The *in vitro* part of the assay follows as testes are sectioned or ovaries are minced (the minced ovary method is described below). The *in vitro* preparations are e.g. whole organ, sections, isolated/cultured cells etc.	Blood samples can be collected during the in life period of the test. Adding different substances (e.g. hCG) can assess the activity of steroidogenic enzymes.	Species sensitivity rat > hamster > rabbit. Test recommended on Sprague-Dawley rats. + Reliable, easy to set up and conduct. +Cytoarchitecture of the testes is preserved and the complete and steroidogenic pathway is present in the testicular section. +Detects inhibitors and stimulants of the stereoidgenic pathway. +highly sensitive method because it includes *in vivo* part. +full metabolic activation capacity.	Not yet standardised.
Ex vivo test Whole-minced ovary culture (Gray et al., 1997)	Ovaries from dosed females are minced and incubated with chemical.	Hormone measurements.	- Risk of diurnal fluctuations in hormone production if necropsy is not performed at the same time. - Unable metabolism of chemicals, and only chemicals acting directly will be detected if exposure starts *in vitro* and not *in vivo*. + More reliable than evaluation of hormones in serum.	

In conclusion, more tests are being performed and developed in this area, however, none of the *in vivo* studies have been thoroughly validated and standardised. Therefore, the list of available tests is constantly increasing/altering and results from individual assays are difficult to compare due to variations of test parameters in the performed tests. The ECT-test of steroidogenesis provides information on the ability of a substance to be an ED acting on the production of steroids in adult animals. Positive results in this test indicate that the substance may cause adverse effects in a pubertal assay and/or a two-generation test. The *ex vivo* methods are used to assess substances for altering steroid production and secretion (EPA 2002). In general these *ex vivo* methods are considered

to be more relevant than in vitro screens, because the method combines *in vivo* exposure with an *in vitro* assessment of the effects. The limiting factor on the sensitivity of the test is the inclusion of the *in vitro* assessment that does not account for the HPA. A positive result in the *ex vivo* studies indicates an effect *in vivo* and therefore would give some basis for concern.

2.3.4 Assays to detect chemicals with effects on the thyroid

The thyroid is an important endocrine organ that plays an essential role in normal development of the embryo and the postnatal maturation of animals. For example, foetal testis development, brain maturation and behavioural functions are highly dependent on thyroid-produced iodothyronines T3 and T4. The major role of T3 and T4 in adults is to regulate energy metabolism and endocrine function. On a cellular level iodothyronines play an important role in synthesis of membrane proteins, enzymes and hormones as well as they incorporate and accumulate iodine in large quantities. In humans T4 is produced in the thyroid and is converted to T3 in the liver and kidneys only whereas in rats the conversion also takes place in the thyroid. Both iodothyronines are incorporated in thyroglobin, which is the protein that synthesise and stores these two hormones. From here both T3 and T4 can rapidly be secreted by enzymatic cleavage. The function of T3 is believed to be major compared with T4 because nuclear receptors bind T3 ten times stronger than T4. Binding to nuclear receptors stimulated the synthesis of membrane proteins, enzymes and hormones. The also stimulate thermogenesis and facilitates the utilisation of glucose and lipolysis. Furthermore, T3 is necessary for maintenance of normal gonadotropin production by the pituitary (Nordic Council of Ministers, 2002).

There is concern because some chemicals exert disruptive effects on thyroid function in at least wild life species, and similar effects might occur in humans. Wildlife studies has revealed that the thyroid is enlarged in Salmon from PCB-contaminated waters and laboratory rodents fed such fish get enlarged thyroid glands too (Hormonally active agents in the environment- Nat. Aca. Press, 1999). A substance that inhibits thyroid function directly is called a goitrogen. Thyroid dysfunction can lead to abnormal development and altered growth patterns in mammals (EPA, 1998). However, the thyroids also play a role in the reproductive system, e.g. there are receptors for T3 in Sertoli cells in the testes (Stoker et al., 2000).

Table 6. Assays for thyroid activity.

Test	Design	Endpoints	Advantages/Limitations	Guidelines
Frog metamorphosis assay (Gray et al., 2002; EPA, EDSTAC final report 1998).	Species: X. laevis Initiated at stage 60, onset of thyroid function. Exposure to chemical for 14 days from day 50 to 64.	Tail resorption: Agonists enhancing resorption and antagonists inhibiting this metamorphosis.	+ Chemistry of the thyroid hormone and the endocrine mechanisms of thyroid hormones are similar among vertebrates. - Not the most Thyroid sensitive period of development. - Large individual differences in the rate of metamorphosis. - Tail tissue is the least sensitive tissue, not specific to alterations in the thyroid system. Chemicals can induce developmental delay. - Not ideal for detection of anti-thyroid effects -Tail resorption is also regulated by corticoids, oestrogens and prolactin.	Recommended by EDSTAC for initial screening of chemicals, is however under evaluation by EPA.

To test whether a chemical disrupts the hypothalamus-pituitary-thyroid axis (HPT-axis), a limited number of screening assays are available. EDSTAC recommends the following *in vivo* assays: The frog metamorphosis assay, The adult male assay (to detect anti-androgenic effects, full thyroid and reproductive hormone screen) and the Japanese quail has been proposed as an in vivo model for action on TH-binding proteins (Ishihara et al., 2003).

None of the tests have yet been validated for guideline purposes, and there is only one vertebrate-based assay currently being evaluated namely the "frog metamorphosis assay" (Gray et al., 2002, EPA), see table 6. This level 3 includes the frog metamorphosis assay although ongoing validation activities concerning this assay in the OECD TGP do not address human health. Tests in mammals are listed and described on level 4.

The frog (tadpole) metamorphosis assay, which is meant to provide indication of mammalian thyroid effects, has been designed to detect EDCs that affect the vertebrate thyroid system. Interference with endocrine function of thyroid is in this test observed as disturbance of tail resorption- agonists enhancing resorption and antagonists inhibiting this metamorphosis (Gray et al., 2002). These endpoints are not specifically related to effects on humans. However, they might reflect or indicate that the chemical might also interfere with thyroid function in humans and other mammals because the endocrine mechanisms of thyroid hormones are similar among vertebrates.

In conclusion, it is at this moment uncertain to what extent the frog metamorphosis assay can be used in screening for effects in humans.

2.3.5 Conclusions at level 3

The assays at level 3 provide an *in vivo* screening of potential endocrine disrupting activity of a substance. Except for the frog metamorphosis assay, the assays at level three provides information about the potency of the compound *in vivo*. Furthermore, the outcome of the assays indicates potential for adverse effects in the reproductive-developmental studies at level 5. At present, it is uncertain to what extent the frog metamorphosis assay can be used for screening in relation to effects on humans.

2.4 Level 4, male and female pubertal assays, intact male assay, and enhanced OECD TG 407 (28-day toxicity study) providing in vivo data about multiple endocrine mechanisms and effects

2.4.1 Introduction

Level 4 is a level between short-term *in vivo* screenings test like the Uterotrophic and the Hershberger assay and the generation studies. The assays can be used for detecting a broad spectrum of endocrine disrupters. In addition, they may provide knowledge of the mode of action that can identify the specific types of endocrine activity. Table 7 shows an overview of the assays at level 4, while a more detailed summary of the endpoints in each of the assays is given in the Appendix.

This level includes the male and female pubertal assay, the intact male assay and the OECD TG 407. Endocrine Disrupter Screening and Testing Advisory Committee (EDSTAC) has recommended a Tier 1 screening battery and two alternate screening

batteries for detecting endocrine-acting compounds in vivo. Besides from the in vitro assays these screening approaches include either 1) the female pubertal assay together with the Uterotrophic (3-day exposure) and the Hershberger assay, 2) the male pubertal assay together with the Uterotrophic assay (3-day exposure) or 3) the intact adult male assay together with the Uterotrophic assay (5-day exposure). The OECD enhanced TG 407 protocol includes endpoints in order to detect chemicals affecting oestrogenic, androgenic and thyroid responsive functions. This assay has both male and female rats represented and compared to the other assays at level 4 the enhanced TG 407 is currently being validated within OECD.

Table 7. Overview of Level 4 tests

Pubertal female assay	Postnatal day 22-42. Daily dosing by oral gavage. n=15 in each of minimum two treatment groups.	Growth, serum T_4 and TSH, age at vaginal opening, vaginal cytology, ovarian and uterus weight and histology, weight of liver, kidney, pituitary, adrenals. Optional endpoints: T_3 estradiol and prolactin.	+Sensitive to modulators of the HPG-axis and the thyroid. +Endpoints the same as in the generation studies. - Do not detect 5α-reductase and some aromatase inhibitors. - Growth and nutritional status influence vaginal opening.	Recommended by EDSTAC[a] to be part of a Tier 1 *in vivo* screening battery together with the Uterotrophic and the Hershberger assay.
Pubertal male assay	Postnatal day 23-54. Daily dosing by oral gavage. n=15 in each of minimum two treatment groups.	Growth, age at balanopreputial separation, serum T_4 and TSH, weight of seminal vesicles, levator ani/bulbocavernosus muscle, and ventral prostate. Thyroid, testis and epididymal weight and histology. Optional endpoints: liver, kidney, adrenal, pituitary weights, serum testosterone, estradiol, LH, T_3, and prolactin. *Ex vivo* testis and pituitary hormone production, hypothalamic neurotransmitter levels, sperm motility and concentration in cauda epididymis.	+Sensitive to modulators of the HPG-axis and the thyroid. +Endpoints the same as in the generations studies. -Do not detect all aromatase inhibitors. -Growth and nutritional status influence preputial separation.	EDSTAC has also considered an alternative Tier 1 *in vivo* screening battery, which includes the pubertal male assay and the Uterotrophic assay.
Intact male assay	Postnatal day 70-90. Originally, daily i.p. dosing for 15 days was proposed. Recent studies have demonstrated that oral dosing is within the same range of sensitivity as i.p. dosing. n=15 in each of minimum two treatment groups.	Weight of liver, seminal vesicle and ventral prostate. Thyroid, testes and epididymides weight and histology. Serum levels of T_3, T_4, TSH, oestradiol, prolactin, testosterone, dihydrotestosterone, LH, FSH.	+Characterization of mode of action. +Include hormone levels as mandatory endpoints. +Sensitive to modulators of the HPG-axis and the thyroid. -Controversy about the sensitivity in young adults compared to the pubertal assays.	Considered by EDSTAC as another alternative Tier 1 *in vivo* screening battery together with the Uterotrophic assay.
Enhanced TG 407	28 day s.c. or p.o. administration to intact 7 weeks old female and male rats (n=10/group, 5 male and 5 female/dose) At least 3 treatment group and 1 vehicle group, the lowest dose level should demonstrate a NOAEL. A satellite group of 10 animals kept for delayed effects/recovery for at least 14 days.	*Hormone analysis (T_3, T_4 and TSH). Female oestrous cycle. Sperm analysis (epididymal sperm counts and morphology) at necropsy. Histopathology restricted to target organs, endocrine sensitive organs, liver, kidneys and adrenals. Growth. Haematology and clinical biochemistry.	+Sensitive to alteration of HPG-axis and thyroid function. + Uses intact animal (without surgical manipulation) -Uses live animal, less sensitive than the Hershberger assay in detecting AR ligands. -The exposure period does not cover one full spermatogenic cycle in rats. - Uses only 5 animals/sex/dose group.	In preparation: The first phase of the validation of the enhanced TG 407 is completed and the validation of the second phase is in progress within OECD.

* The endpoints listed here are the endocrine-related endpoints that extend the existent OECD 407 protocol.
[a] EDSTAC Endocrine Disrupter Screening and Testing Advisory Committee

2.4.2 Description of tests

Pubertal female assay
The purpose of this protocol is to outline procedures to quantify the effects of endocrine disrupters on pubertal development and thyroid function in the intact juvenile/peripubertal female rat. The test is designed to detect ER agonists and antagonists, thyroid modulators, steroid biosynthesis inhibitors, and compounds that alter the hypothalamic-pituitary-gonadal axis. The test is not validated.

Pubertal male assay
The purpose of this protocol is to outline procedures to quantify the effects of endocrine disrupters on pubertal development and thyroid function in the intact juvenile/peripubertal male rat. By combining the endpoints several antiandrogenic as well as oestrogenic compounds have been identified. The test is not validated.

Intact male assay
This assay can unlike the pubertal male and female assay detect aromatase inhibitors. It is designed to run in parallel with the Uterotrophic assay and the in vitro receptor binding and/or transcriptional activation assays.

This assay identifies a broad spectrum of endocrine disrupters and aids in the characterisation of their mode of action. Using the intraperitoneal route several chemicals have been identified by this protocol as endocrine disrupters including 17β-estradiol, coumestrol, ICI-182,780, testosterone, flutamide p,p'-DDE (only in Long Evans rats but not in CD rats), vinclozolin, progesterone, mefepristone (Progesterone receptor antagonist), haloperidol (D2-receptor antagonist), reserpine (dopamine depletory) phenobartital, propylthiouracil, finasteride, ketoconazole, anastrozole (O'Connor et al., 2002a).

The chemicals, flutamide (a strong AR antagonist), p,p'-DDE (a weak AR antagonist), ketoconazole (a testosterone biosynthesis inhibitor), and the two thyroid modulators phenobarbital and propylthiouracil have besides from being tested by the intraperitoneal route also been dosed by oral gavage. Overall, the conclusions of the studies were that the sensitivity was within similar magnitude between the two different routes of administration (O'Connor et al., 2002b and c). The test is not formally validated in the OECD TGP.

Enhanced OECD TG 407
The current OECD 407 test guideline is enhanced with several new endocrine related endpoints as a screening method for detection of chemicals with endocrine mediated effect (table 7, OECD 2002c). With the additional endpoints, this assay may identify chemicals affecting the oestrogen, androgen or thyroid system.

Enhancement of TG 407 allows for a more complete assessment of the endocrine system compared to the two short-term in vivo screenings assay at level 3, because TG 407 uses intact animals and has multiple endpoints.

Assessment of the sensitivity and reliability of the enhanced TG 407 protocol to identify endocrine disrupting chemicals and their putative target tissues is still ongoing within the OECD TGP. Most of the endpoints, for instance, oestrous cycle and sperm parameters are also included in guidelines for other tests.

2.4.3 Endpoints

Growth

Nutritional status and body weight of mammals have been recognised as having effects on reproduction and puberty (Goldman et al., 2000). Toxicants administered during puberty can result in decreases or increases in body weight, thereby affecting sexual maturation. However, oestrogen administered prior to puberty results in an advance in puberty in females, despite the decrease in body weight (Goldman et al., 2000). Methoxychlor, a weak oestrogen and antiandrogen, dosed to juvenile male and female rats decreased the body weight in both sexes, delayed preputial separation in the males and accelerated vaginal opening in the females (Gray et al., 1989). The female pubertal assay is more sensitive to oestrogenic compounds compared to the male assays (Stoker et al., 2000). Body weight changes are considered adverse, but it can be difficult to elucidate, whether the effect is endocrine related, a general toxic effect, or caused by changes in numerous other factors involved in the relationship between body weight and sexual maturation.

Thyroid status

The thyroid hormones are essential for regulation of normal growth and development in young animals. The ability of pharmaceuticals and environmental chemicals to disrupt the thyroid system by altering biosynthesis, secretion or metabolism has been studied in both human and animal models. Several environmental chemicals have been shown to decrease T3 and T4 and elevate TSH. Induced alteration of thyroid hormones during puberty may cause permanent effects regarding growth and development.

Serum T4 and TSH analysis are included in order to detect thyroid-active agents in both the pubertal male and female assay as well as in the intact male assay. However, some compounds with anti-thyroidal properties will not be detected, because of T4 or TSH compensation. A combination of hormone measurements with histopathology of the thyroid gland (epithelial cell height and area of thyroid follicle colloid) are required endpoints in the male and female pubertal assay proposed by EDSTAC, while measuring of T3 levels is optional. Chemicals with the ability to alter thyroid function in rodents often induce decreased serum levels of T3 and T4 and increased serum TSH levels. The sustained release of TSH results in follicular cell hypertrophy/hyperplasia. Male rats are more sensitive than female rats t identify thyroid toxicants due to the higher circulating levels of TSH and thus the female pubertal assay may have a lower sensitivity for detecting thyroid-modulating compounds. When body weight decrements are 15% or greater confounding of the data can occur (O'Connor et al., 1999). Interestingly, antiandrogens like p,p'-DDE, DBP, linuron, vinclozolin and prochloraz also affect thyroid endpoints (O'Connor et al., 2002b; Vinggaard et al., 2002).

Mechanisms of effects related to thyroid gland function are listed below and include both a direct and indirect toxic effects. These are summarised in TemaNord (Nordic Council of Ministers, 2002) and the reader is referred to this report for further information about thyroid related mechanisms and effects.

Direct effects:

- Inhibition of iodide uptake and transport into the thyroid
- Inhibition of thyroid peroxidase
- Inhibition of iodothhyronine secretion

- Direct damage of follicular cells, which leads to inflammation, and degeneration of the thyroid gland
- Damage of follicular cell DNA leading to thyroid tumour development
- Disruption of iodothyronine biosynthesis

Indirect effects:

- Changes in iodothyronine plasma transport capacity, which leads to displacement of T4 from TTR
- Increased iodothyronine excretion rate
- Inhibition of T4 conversion to T3 by iodothyronine deiodinases
- Interference of intestine reabsorbation of T3 and T4 leading to increase in hormone clearance
- Interference with the hypothalamic-pituitary gland.

Additional hormone measures

In the enhanced TG 407 and pubertal male assay, serum LH, testosterone, oestradiol, T3 and prolactin levels were listed as optional. The primary use of these measures would be to assist in clarifying the mode of action of the test compound. Results from the enhanced TG 407 phase one studies show high animal-to-animal variability in additional hormone levels (LH, FSH, testosterone, prolactin). Measurements of these hormones are therefore optional (OECD, 2003c). Despite the huge inter-animal variability these hormones are investigated in the intact male assay in order to characterize the mode of action early in the testing procedure. According to O'Connor et al. (2002a), mature animals are more sensitive to alterations in serum hormone levels compared to immature animals exemplified in a study where phenobarbital-induced effects on testosterone and DHT were not observed in the pubertal male assay as they were in the intact male assay. There are, however, some conflicting data suggesting that the sensitivity in peripubertal males compared to older males is depending on the endpoints tested as well as mode of action (Stoker et al., 2000).

When pituitary hormones are measured, animal stressors like euthanasia in CO2 should be avoided, as these hormones can be secondary induced by stress. Preferably the animals should be quickly decapitated without anaesthesia (maximum 15 seconds) after leaving their cage (Goldman et al., 2000; Stoker et al., 2000).

Oestrous cycle

The female rat has a regular cycling period in common with e.g. women. The female rat displays a four-five days cycling pattern, shown by different patterns in vaginal cytology. This is a direct reflection of the change in circulating concentration of oestradiol and progesterone. Environmental chemicals may induce alteration in the cycle length and thereby in the reproductive function. A study by Yamasaki (et al. 2002a) shows that bisphenol-A is capable of prolonging diestrous at 600 mg/kg/day. Additionally, abnormal levels of thyroid hormone (e.g. in hypothyroidism) can alter oestrous cyclicity in the female rat (Gray et al., 1997). In general, a statistically significant change in the oestrous cycle is considered to be adverse.

Vaginal opening

Environmental oestrogens, aromatizable androgens and dopamine antagonists are able to advance vaginal opening (Goldman et al., 2000). Higher brain centres play a critical

role in puberty onset by controlling hypothalamic output leading to secondary control of anterior pituitary function and ovarian maturation. Therefore effects on these areas of the brain may affect vaginal opening and first ovulation. Examples of delayed vaginal opening are observed after dosing with morphine or fentanyl, mu-opioid receptor agonists. Vaginal opening may also be influenced by body weight, feed intake, diet composition, nutrient deficiencies, and agents influencing the hypothalamic-pituitary-gonadal axis. The numerous factors contributing to pubertal onset suggest that this endpoint should be included in a battery of endpoints.

Balano-preputial separation
Preputial separation is a noninvasive indicator of the androgen status of the pubertal rat. However, a delay in preputial separation is not only induced by antiandrogens and oestrogens, as compounds influencing alterations of prolactin, growth hormone, gonadotropin secretion or hypothalamic lesions also alter the rate of pubertal maturation in weanling rats. Also environmental factors, like nutritional status and weight loss can influence age at preputial separation. Therefore, it should be considered whether the effects are related to endocrine disruption. The stadium of separation should be recorded as well as any additional observations, such as a persistent thread between the glans and the prepuce.

Sperm analysis
In the assessment of sperm quality, sperm count and sperm morphology are mandatory endpoints. Results from the OECD phase one study show that sperm motility is a less sensitive endpoint due to a high animal-to-animal variability. Sperm motility is therefore not included in the OECD phase two study. The sample method in sperm analysis is crucial in reducing variability. The sampling method is not being standardised in the present protocol. Alteration in sperm quality is often a result of testicular lesions. Several environmental chemicals have shown to cause a decrease in sperm quality. In general, a statistically significant decrease in sperm quality will be considered adverse and indicate potential effects in human.

Weight of reproductive organs
The epididymides, ventral prostate, seminal vesicles and levator ani/bulbocavernosus muscle are all androgen sensitive tissues that generally show reduced size following exposure to antiandrogens including compounds affecting testosterone biosynthesis inhibitors and 5α-reductase. However, depending on the mechanism of action, these organs may respond differentially. For example the weight of ventral prostate and seminal vesicles can either be unchanged or increased after dosing with an aromatase inhibitor (O'Connor et al., 2002a). This information coupled with hormone analysis can be valuable in identifying mode of action in the screening assays.

In general, the weight of uterus and ovaries are either unchanged or decreased regardless of mode of action in the female pubertal assay (O'Connor et al., 2002a).

Histology
Severity of effects induced by endocrine disrupters can often be determined by histopathology of various organs e.g. vaginal cornification, epithelial cell height in the uterus and thyroid, testes atrophy, Leydig cell hyperplasia, spermatid retention. These effects can contribute to information about mode of action of a compound. Thyroid gland histopathology has been determined to be the most reliable parameter for the detection of compounds that affect thyroid function (O'Connor et al., 1999).

2.4.4 Discussion

Pubertal male and female assay

These assays have a broad application, are simple to conduct and have endpoints in common with the multi-generation tests. There is no surgical manipulation of the animals, and the endpoints are sensitive and tend to be rather robust in the detection of endpoints related to endocrine disrupters (Stoker et al., 2000).

According to O'Connor et al. (2002a), the female pubertal assay detects endocrine disrupters including ER agonists and antagonists, potent thyroid agents and the aromatase inhibitor fadrozole, steroid biosynthesis inhibitors, aromatizable androgens and prolactin modulators. Data from finasteride studies indicate that this assay does not detect 5α-reductase inhibitors (Marty et al., 1999). Furthermore, this assay does not detect δ–testolactone, a moderately specific aromatase inhibitor, using a concentration that induces antiandrogenic effects in pubertal males (Marty et al., 1999). Hence, δ–testolactone possibly possesses other modes of action besides from its aromatase inhibiting effects.

Concerning the male pubertal assay it seems sensitive to alterations of the HPG axis and steroidogenesis including the aromatase inhibitor (anastrozole) as well as androgen receptor agonists and antagonists. The mechanism of action may not be apparent without applying the proposed optional endpoint (see table 7).

Compared to the female pubertal assay, this male assay seems less sensitive to xenooestrogens. Additionally, the aromatase inhibitor, fadrozole, was not detected in this assay probably because the role of oestrogens in male pubertal onset is limited (O'Connor et al., 2002a). Fadrozole is detected both in the pubertal female and the intact male assay but this may be explained by the use of a lower dose in this male assay (6 mg/kg bw) compared to the dose used in the intact male assay (25 mg/kg bw) (O'Connor et al., 2002a).

The mechanism of sexual development during puberty is similar among mammals including humans. Therefore, tests on laboratory animals can be used in human risk assessment. There has, however, been some controversy about the necessity for dosing rats during the full 30 days and an alternative protocol including only 14 days of dosing has been suggested (Ashby and Lefevre, 1997). According to Stoker et al. (2000), two weeks is sufficient for some chemicals to induce changes of the thyroid gland, but a 30-day exposure is needed to detect weak-acting chemicals that target the thyroid gland.

Many chemicals, for example PCB congeners and brominated flame retardants, may affect thyroid homeostasis and many often uninvestigated mechanisms may be involved. Therefore, the histopathology of the thyroid becomes crucial to investigate in relation to manifestation of adverse effects. The sensitivity of the pubertal male and female assays may be improved by adding the optional hormone investigations to the protocols. Furthermore, this would aid in characterising the mode of action of the investigated endocrine disrupters.

Intact male assay

This 15-day screening assay can identify endocrine disrupters with several different modes of action. When changes in organ weights are coupled with the serum hormone data for an unknown compound, the mode of action can be characterised. For example an AR antagonist, like flutamide, competes with testosterone and DHT for binding to the androgen receptor thereby blocking the recognition of testosterone and DHT.

Consequently, the weight of accessory reproductive glands is decreased and serum testosterone and LH concentration are increased. A steroid biosynthesis inhibitor like ketoconazole acts directly at the testis and decreases testosterone production. The weight of accessory reproductive glands decreases, serum testosterone levels decreases and secondary to this LH levels increases. Thus, the "fingerprints" of the compounds in this assay can be used to distinguish between these two different modes of actions. By identifying the potential mode of action, critical endpoints can be included in reproductive toxicity studies. Generally, the antiandrogens investigated in O'Connor et al., (2002 b) caused changes in hormone levels at lower dose levels compared to the doses, where organ weight changes were observed. By including hormone levels to provide information about LO(A)ELs, the sensitivity of the intact male assay will most probably become comparable to other in vivo mammalian screening tests at level 3 and level 4.

Enhanced TG 407

The intention of the enhanced TG 407 is to improve the sensitivity of the existent 28-days toxicity protocol to identify endocrine disrupting chemicals. The validation of the enhancement of TG 407 is still an ongoing process within the OECD. The data obtained so far indicate inconsistencies in some of the results among the participating laboratories. Differences in methodologies and sampling amongst the laboratories appear to be the main problem. An example is for instance differences in sperm number and sperm morphology methodologies, where differences in the characteristics and terminology for abnormal sperm exist among laboratories.

The dosing period in the protocol does not cover one full spermatogenic cycle in rats. Spermatogenesis in the rat takes 56 days during which primitive spermatogonia give rise to highly differentiated spermatids. The spermatogenic cycle is divided into series of morphological identifiable stages where four generations of different developed cell types are represented. A detailed histopathological examination of the different stages is a sensitive method for detecting disturbances in spermatogenesis. Therefore, in order to make a thorough assessment of effects on the spermatogenesis it should, after our opinion, be considered to include staging in the protocol. By including spermatogenic staging in the histopathological examination of testicular toxicity the sensitivity of regulatory guideline studies can be improved in general (Creazy 1997).

Five animals per sex per dose group (total of 10 animals/dose group) are used in the phase two protocol. This number could very well be insufficient to achieve a reliable result due to animal-to-animal variation. Therefore as an option in the phase-two-validation study, 10 animals/sex/dose group are examined in a few laboratories. The result from these studies will be used to determine whether 5 animals per sex give a reasonable sensitivity of the protocol.

The enhanced TG 407 is from the OECD point of view not intended to be an ultimate test. The assay is not intended to provide an all-encompassing profile of the endocrine action. A negative result in the enhanced TG 407 would however imply the substance to be of low concern for endocrine mediated effects. A positive finding could be addressed in subsequent studies for further definition of effects and hazards (OECD draft 2003c). According to our evaluation, effects of a substance in the enhanced TG 407 are to be considered as adverse. Therefore, a substance causing this effect should be considered to have a potential for causing effects in humans.

2.4.5 Conclusions at level 4

The assays at level 4 provide a thorough assessment of the potential endocrine disrupting effects of a substance in pubertal and young adults. Furthermore, level 4 provides information about the potency of a compound to be investigated at level 5. Effects on various endpoints included in the assays can either be considered as adverse or represent an effect on a mechanism relevant for humans e.g. changes in hormone levels. Therefore, these assays can be used to provide NO(A)ELs/LO(A)ELs to be used in human risk assessment until further studies are available. The intact male assay and the TG 407 may be more capable for detecting aromatase inhibitors and compounds affecting the steroid synthesis compared to the pubertal male assay. On the other hand, the two assays in intact young males may be less sensitive compared to the Uterotrophic and the Hershberger assay as well as the male and female pubertal assay.

2.5 Level 5, *in vivo* assays providing adverse effect data from endocrine and other mechanisms

2.5.1 Introduction

The box at level 5 in the OECD Conceptual Framework includes "*in vivo* assays providing adverse effect data from endocrine & other mechanisms" and the tests listed are: one-generation assay (TG 415 enhanced), two-generation assay (TG 416 enhanced), reproductive screening test (421 enhanced) and combined 28 day/reproduction screening test (TG 422 enhanced). A note for each assay states that potential enhancement will be considered by the OECD Working Group on testing in mammalians.

However, other *in vivo* assays for reproductive toxicity may be considered relevant for this box such as developmental neurotoxicity tests (draft TG 426) and prenatal developmental toxicity tests (TG 414) and consequently these tests are included in the following.

For each of these assays, the following will include a short description of each test, consider the end points included and their reliability and relevance for humans as well as gaps/deficiencies and give proposals for enhancements. Table 8 gives an overview of the design, endpoints and advantages/limitations in relation to EDC effects, while a more detailed summary of the endpoints assessed in each of the tests are given in the Appendix.

2.5.2 Reproductive toxicity tests

One- and two-generation studies (OECD TG 415 and 416)
The purpose of generation studies is to examine successive generations to identify possible effects of a substance on fertility of male and female animals, pre-, peri-, and postnatal effects on the ovum, foetus, and progeny, including teratogenic and mutagenic effects, and peri- and postnatal effects on the mother.

The preferred species are the rat and the mouse. Other species may be used if relevant, for example in case of differences in toxicokinetics between the preferred species and man, to clarify ambiguous results or to further study observed effects.

The number of animals per group should be sufficient to yield about 20 pregnant animals at or near term. The chemical is administered over at least one spermatogenic cycle and the last stages of oocyte maturation before the parent generation animals are mated. The exposure of the females is continued throughout the mating period and the gestation up to weaning of the last generation, i.e. F1 in the one-generation study and F2 in the two-generation study. At least three treatment groups and a control group should be used. Ideally, unless limited by the physico-chemical nature or biological effects of the test substance, the highest dose level should induce some toxicity but not mortality in the parental animals. The low dose should ideally not induce any observable adverse effects on the parents and offspring.

Table 8. Overview of *in vivo* OECD guideline tests for reproductive toxicity testing

Test	Design	Endpoints	Advantages/limitations
OECD TG 416 Two-Generation study	Exposure before mating for at least one spermatogenic cycle until weaning of 2nd generation. 3 dose levels plus control N = 20 pregnant females.	Fertility. Oestrus cyclicity and sperm quality. Growth, development and viability. Anogenital distance if triggered. Sexual maturation. Histopathology and weight of reproductive organs, brain and target organs.	+exposure covers all sensitive periods. +effect assessment in F1 and F2. +includes assessment of semen quality and oestrus cyclist. -anogenital distance only assessed in F2 if triggered. -areola/nipple retention is not assessed. -malformations of reproductive organs only investigated in 1 per sex per litter.
OECD TG 415 One-Generation Study	Exposure before mating for at least one spermatogenic cycle until weaning of 1st generation. 3 dose levels plus control N = 20 pregnant females.	Fertility. Growth, development and viability. Histopathology and weight of reproductive organs, brain and target organs.	+ exposure covers most of the sensitive periods. - no exposure from weaning to sexual maturation. - not updated to include similar endpoints as the two-generation study.
OECD TG 414 Prenatal Developmental Toxicity Study (Teratology study)	At least from implantation to one or two days before expected birth. 3 dose levels plus control N = 20 pregnant females.	Implantation, resorptions. Foetal growth. Morphological variations and malformations.	+ malformations are assessed in all foetuses. - the dosing period includes only the prenatal period. - the effects assessment includes only effects in foetuses.
OECD TG 426 Developmental Neurotoxicity Study (draft)	At least from implantation throughout lactation (PND 20). 3 dose levels plus control. N = 20 recommended, less than 16 not appropriate.	Birth and pregnancy length. Growth, development and viability. Physical and functional matura-tion. Behavioural changes due to CNS and PNS effects. Brain weights and neuropathology.	+ exposure covers most of the sensitive periods. - no exposure before mating and from weaning to sexual maturation. - mating and nursing behaviour is not assessed.
OECD TG 421 and 422 Reproduction/ Developmental toxicity screening test	From 2 weeks prior to mating until at least day 4 post-natally. 3 dose levels plus control. N = 8-10 pregnant females.	Fertility. Pregnancy length and birth. Foetal and pup growth and survi-val until day 3.	+ short-term test. - limited exposure period. - limited number of endpoints. - limited sensitivity due to number of animals.

In the updated OECD TG 416 two-generation reproductive toxicity study, assessment of effects on sperm quality and oestrus cyclicity in parental animals and F1 has been included. In addition, the updated guideline suggests an improved assessment of the

offspring i.e. the recording of developmental milestones including some behavioural parameters and histopathology of reproductive organs, brain and identified target organs.

The one-generation study has not yet been updated.

Prenatal developmental toxicity study (OECD TG 414)

The prenatal developmental toxicity study is the in vivo method for studying embryo-foetal toxicity as a consequence of exposure during pregnancy. The test was originally designed as a teratology test to detect malformations, and in the past there was a tendency to consider only malformations or malformations and death as relevant endpoints. Today it is assumed that all of developmental toxicity e.g. death, structural abnormalities, growth alterations, and functional deficits are of concern, and consequently, the title of the OECD TG 414 has been changed to "Prenatal Developmental Toxicity Study".

The preferred species include rodents e.g. the rat and mouse and non-rodents e.g. the rabbit. Other species may be used if relevant, for example in case of differences in toxicokinetics between the preferred species and man, to clarify ambiguous results or to further study observed effects.

Young mature virgin females are artificially inseminated or mated with males. The time of mating is established by observation of mating (e.g. rabbits), identification of a plug (mixture of sperm and cellular material from the vagina of rats and mice), vaginal smear (in rats) or by noting the time of insemination (e.g. for pigs and rabbits). Three dose levels and a control group are used in order to establish a dose-effect relationship. The group size is 20 pregnant animals for rats and mice, and at least 16 for the rabbit. The pregnant female rats were up to recently exposed at least during the period of organogenesis, i.e. between gestation day 6 when implantation occurs, and gestation day 15. The corresponding periods for mice and rabbits are days 6-15 and days 6-18, respectively. This period has been found to be the most sensitive to the induction of structural, anatomical malformations. However, the development of e.g. the reproductive organs and the brain continues after gestation day 15 and consequently malformations of such organs will not be discovered if exposure is stopped on day 15. In the updated OECD TG 414, the dosing period shall cover the period from implantation to scheduled caesarean section. If preliminary studies, when available, do not indicate a high potential for pre-implantation loss, treatment may be extended to include the entire period of gestation, from mating to the day prior to scheduled kill.

The animals are observed daily for clinical changes. Body weight is recorded and food consumption is recorded throughout the gestation. The day before anticipated birth, the uterus is removed by caesarean section, and the uterus and the foetuses are examined. The dam is examined macroscopically for any structural abnormalities or pathological changes which may have influenced the pregnancy. If dosing is initiated before or at the time of implantation, the pre-implantation loss, i.e. the number of embryos lost prior to implantation is evaluated.

The total number of implantations, i.e. living embryos, dead embryos and resorptions (embryos that die early and are reassimilated, corresponding to early abortions in humans) are noted. The degree of resorption is recorded in order to establish the time of death of the embryo during the pregnancy.

The foetuses are sexed, weighed and examined for gross malformations. Retarded growth and effects on visceral and skeletal development are evaluated including the degree of ossification of the bones.

Developmental Neurotoxicity Study (OECD draft TG 426)
Developmental neurotoxicity studies are designed to develop data on the potential functional and morphological hazards to the nervous system arising in the offspring from exposure of the mother during pregnancy and lactation. These studies investigate changes in behaviour due to effects on the central nervous system (CNS) and the peripheral nervous system (PNS). As behaviour also may be affected by the function of other organs such as liver, kidneys and the endocrine system, toxic effects on these organs in offspring may also be reflected in general changes in behaviour. No single test is able to reflect the entire complex and intricate function of behaviour. For testing behaviour, therefore, a range of parameters, a "test battery", is used to identify changes in individual functions.

The most frequently used animal species are the rat and the mouse. The early guidelines generally recommended groups of 20 animals with dosing from day 15 of gestation to day 21 post gestation i.e. spanning foetogenesis and the entire lactation period. However, the recommended dosing period does not cover all relevant developmental phases since the CNS is also susceptible to abnormal development during the period of organogenesis. Consequently, dosing is often started earlier, for example on day one or day six of the pregnancy. In the draft OECD Guideline 426 the dosing period includes the entire gestation and lactation period. This period covers most of the relevant developmental phases, but not the entire period of brain development as the brain does not attain an approximate adult stage until the age of six weeks, corresponding to 12-15 years of age in humans.

After birth, the number of progeny is recorded and the litters may be adjusted so that each contains the same number of pups. The evaluation of the offspring consists of observations to detect gross neurological and behavioural abnormalities, assessment of physical development, reflex ontogeny, motor activity, motor and sensory function, and learning and memory; and evaluation of brain weights and neuropathology during postnatal development and adulthood.

The draft OECD Guideline 426 is designed to be performed as a separate study, however, the observations and measurements can also be incorporated into e.g. a two-generation study.

Reproduction/Developmental Toxicity Screening Tests (OECD TG 421 and 422)
The purpose of the Reproduction/Developmental Toxicity Screening Test's is to generate limited information concerning the effects of a test substance on male and female reproductive performance such as gonadal function, mating behaviour, conception, development of conceptus and parturition. It is not suggested as an alternative to or as a replacement for the existing test guidelines for generation and developmental toxicity studies, but rather as an in vivo screening assay.

The dosing of the animals is initiated two weeks prior to mating and continued until the end of the study on postnatal day four. The number of animals per group is at least 10 animals of each sex and is expected to provide at least eight pregnant females per group.

Effects on fertility and birth are registered. Live pups are counted and sexed and litters weighed on days one and four postpartum. The parameters include among others a

detailed histological examination of the ovaries, testes and epididymides of at least the highest dosed and the control animals.

The tests do not provide complete information on all of those aspects of reproduction and development covered in the OECD one- and two-generation studies, the developmental neurotoxicity study and the prenatal developmental toxicity study. In particular, it offers only limited means of detecting postnatal manifestations of prenatal exposure or effect induced during postnatal exposure.

The value of a negative study is more limited than data from generation and developmental toxicity studies due to the lower number of animals per group, the shorter period of exposure as well as the limited number of endpoints measured.

2.5.3 Endpoints

Fertility
Assessment of fertility is included in generation studies and in the reproductive screening tests. Assessment of fertility in exposed offspring is only possible in the two-generation study.

Fertility is generally expressed as indices that are ratios derived from the data collected. For example, the mating index is used as a measure of the male's or female's ability to mate and is defined as: number of animals with confirmed mating/the total number of animals cohabitated. The fertility index is a measure of the ability to achieve pregnancy and is defined as: number of males impregnating a female or number of pregnant females/total number of animals cohabitated. The effects on fertility may be related either to effects induced before mating or thereafter, but before implantation i.e. pre-implantation losses. The data will not necessarily be similar for studies in which there is more than one mating per generation or more that one mating generation.

The interpretation of fertility data should consider the duration of treatment and the number of animals investigated.

The males are not dosed during the total of the spermatogenic cycle in the reproductive screening tests and the studies use a limited number of animals. Therefore, it is unlikely that reproductive effects are manifested on the fertility data.

Reduced fertility has been found for a number of EDCs, but often at relatively high exposure levels. However, it has to be considered that the rat is the most commonly used experimental animal and that a male rat can generally still produce normal progeny after having its sperm production reduced to 10% of the normal level (Aafjes et al 1980). Thus, fertility data from rat studies alone can be a rather insensitive endpoint. Human males are not supposed to have a similar sperm reserve capacity as rodents and therefore the 2-generation study has recently been updated to include assessment of sperm quality. It is unknown whether female rats compared to women also are less sensitive to effects on fertility, but the 2-generation study has recently also been updated in that aspect by the inclusion of assessment of oestrus cyclicity.

Sperm quality and oestrus cyclicity
These endpoints are included in the revised two-generation study, but at present not in the one-generation study. However, the one-generation study could be updated to include assessment in the parental animals without significant changes of the design. Assessment of these endpoints in exposed offspring is only possible in the two-

generation study design. However, it would be possible to include the endpoints in the one-generation study if the study period were extended to around postnatal day 90 instead of postnatal day 21.

The parameters included for assessment of sperm quality are sperm number, sperm morphology, and sperm motility. Testicular lesions of sufficient magnitude can impact sperm quality, but normal sperm quality is dependent on a number of other factors. Therefore, changes in sperm parameters should not be discounted in the absence of histological lesions. For example, sperm changes could be due to effects on epididymides. In general, a statistically significant change in sperm parameters would be interpreted as indicating a potential effect on human fertility.

Vaginal cytology is evaluated to determine the length and normality of the oestrous cycle in P and F1 females in the two-generation study. The data can provide information on cycle length, persistence of oestrus or dioestrus, and incidence of pseudopregnancy. An effect on oestrous cycle can affect reproductive performance, but this will depend on the nature and the magnitude of the effect. In general, a statistically significant change in the length of the cycle or prolonged oestrus or dioestrus would be considered adverse.

Anogenital distance

In the two-generation study, anogenital distance (AGD) should be measured at postnatal day 0 in F2 pups if triggered by alterations in F1 sex ratio or timing of sexual maturation.

The anogenital distance is longer in males than in females. As there is a relationship between the anogenital distance and size of the pup, the body weight of the pup is normally used as a covariate in the analysis of the data. However, if body weight is significantly influenced by exposure this may actually mask effects on anogenital distance.

In general, a statistically significant change in the anogenital distance that is not related to the size of the pups would be considered adverse.

Several studies have shown that hormones and EDCs may change the anogenital distance. This has especially been shown for antiandrogens, where decreased AGD has been shown for e.g. some phthalates, procymidone and vinclozolin (Mylchrest et al., 1999, Ostby et al., 1999, Gray et al., 1999a).

As AGD is an endocrine sensitive endpoint it is considered relevant to include assessment of AGD in both F1 and F2 as a standard parameter in the 2-generation study without the need for triggering this by letting it be triggered by alterations in sex ratio or timing of sexual maturation. In addition, the parameter could be included in the one-generation study and the reproductive toxicity screening tests, i.e. TG 421 and 422.

Nipple/areola retention

Assessment of nipple or areola (dark area around the nipple bud) retention in male offspring is not included at present in any OECD TG.

In female offspring 12 areolas (dark focal areas) are normally visible around postnatal day 13, while very few or none are visible in male offspring. As the development of fur in the animals makes it difficult (impossible) to see the areolas a few days later it is important to establish the correct time for the assessment in the animals used for the study. Often, only the presence or absence is registered in the males, however, the number of areolas in each male may provide a more sensitive assessment.

In general, a statistically significant increase in the number of nipples/areolas in male pups would be considered adverse. This is especially the case, if it is shown that some of these nipples are persistent, i.e. they are also present in the adult males.

Several studies have shown nipple retention in male offspring after exposure to hormones and some hormonal disrupters, especially anti-androgens (Gray et al., 1999b, Ostby et al., 1999, Mylchreest et al., 1999, Mylchreest et al., 2000). As nipple or areola retention in male offspring is an endocrine sensitive endpoint it is considered relevant to include assessment of this endpoint in both F1 and F2 offspring as a standard parameter.

Sexual maturation
Assessment of the timing of sexual maturation is included as optional endpoints in TG 416 and TG 426.

Assessment of the onset of puberty in females is done by inspection of vaginal opening. In rats, this occurs around postnatal day 30-35. In males, testicular descent or balano-preputial separation, which occurs around postnatal day 21-30 and 42-48, respectively, may be used as indications of puberty. Observation of testicular descent relies on how the animal is handled during inspection and is rather difficult to assess. Balano-preputial separation corresponds to puberty in male rats (Korenbrot et al., 1977) and is the endpoint included in the Developmental Neurotoxicity Study and the two-generation study. Both the age and the body weight of the animal at sexual maturation are to be recorded, as there is a relationship between these end points.

In general, significant changes in the timing of sexual maturation that is not explained by body weight effects would be considered to indicate a potential effect in humans.

Birth weight and postnatal growth
Birth weight and postnatal growth are recorded in all reproductive toxicity guidelines except in the TG 414, where the foetuses are killed the day before expected birth.

It is important in the evaluation of the data to consider variations due to different litter sizes or different sex distribution in the litters. A change in offspring body weight is a sensitive indicator of developmental toxicity, in part because it is a continuous variable. In some cases, weight reduction in offspring may be the only indicator of developmental toxicity in a generation study. While there is always a question remaining as to whether weight reduction is a permanent or transitory effect, little is known about the long-term consequences of short-term foetal or neonatal weight changes. Therefore, weight reduction is normally used to establish the NOAEL.

Histopathology of reproductive organs
Histological examination of reproductive organs is included in the generation studies and the in vivo reproductive toxicity screening studies (i.e. OECD 421 and 422). The two-generation study that specifies dosing of the male for the entire spermatogenic cycle combined with analysis of sperm quality includes less extensive histopathological examination than the studies where the dosing regime is shorter and no sperm analyses are conducted (e.g. OECD 421 and 422).

Thyroid hormones may cause effects on the testes, as T3 has influence on the balance between Sertoli cell proliferation and differentiation in the lactational period of rats. It has been demonstrated that hypothyroidism and decreased levels of T3 in male offspring during lactation, extend the period of Sertoli cell proliferation, while Sertoli cell maturation is delayed (Simorangkir et al., 1995). Male rats exposed to PCB's during lactation have decreased T3 levels, and consequently increased adult testis

weight as well as increased the number of Sertoli cells followed by an increased number of sperm cells (Cooke et al., 1996).

Histopathological findings are generally classified according to qualitative criteria and the data are presented as the number of animals affected within a dose group. There may not be an obvious relationship between histopathological findings and fertility. For short-term studies (i.e. OECD 421 and 422) in which the animals are treated for less than the duration of spermatogenic cycle, an effect on spermatogenesis may not have had adequate time to become evident as reduced sperm counts.

In general, any dose-related significant histopathological finding would be considered to indicate a potential effect in humans.

Malformations of reproductive organs
Examination of structural changes of the reproductive organs is included in all guideline studies except TG 426. Malformations of reproductive organs after exposure to EDCs include for example persistent nipples, hypospadia and cryptorchism in male offspring.

The OECD 414 is designed especially to investigate major malformations and the pregnant animals are killed prior to the expected day of delivery (gestation day 21 in the rat). The guideline specifies that investigations of the foetuses should be with particular attention to the reproductive tract. The techniques used may however not allow that "small" malformations of the reproductive organs are revealed.

In studies performed according to the OECD TG 414 Teratology Study before the guideline was updated, the animals were dosed only during the major organogenesis (i.e gestation days 6-15 in the rats). As important parts of the development of the reproductive organs happens after gestation day 15, the potential for effects on the development of the reproductive organs cannot be assessed in such studies. In the updated OECD TG 414, the exposure period is extended until a few days before birth. However, the development of the reproductive organs continues after birth and some effects may not be present until the animal is sexually mature. Consequently, the OECD 414 is not suitable for detection of some important malformations of the reproductive organs.

Malformations of the reproductive organs such as hypospadia and cryptorchidism can potentially be found in the generation studies since the offspring is investigated until adulthood. However, only 1/sex/litter is selected at weaning to perform the next generation in contrast to the assessment of malformations using the OECD TG 414, where all foetuses are investigated. As malformations normally are rare effects, the generation studies have a very limited sensitivity for detecting such effects, and therefore, malformations in the reproductive organs occurring at low rates may not reach statistical significance. The short-term studies (i.e. TG 421 and 422) include a limited number of litters and the animals are immature when investigated on postnatal day 3 or 4. Therefore, the potential for malformations of reproductive organs can only be assessed to a very limited extent in these studies.

In general, an increased number of malformations are considered to indicate a potential for effects in humans.

Developmental neurotoxicity
Investigations of developmental neurotoxicity, including behavioural studies and detailed investigations of the brain, are performed only in the Developmental Neurotoxicity Study.

Hormones play a central role in central nervous system development including the sexual differentiation of the brain. Studies on hormones and various EDCs have shown that the developing brain can be susceptible. Brain development is affected by thyroid function (Timiras and Nzekwe 1989) and in the neonatal rat, where the central nervous system is rapidly expanding, a large increase in expression of TH receptors has been observed in the brain (Bernal et al., 1985). Studies in humans have documented associations between sub-clinical hypothyroidism in pregnant women and neurological development of the offspring, e.g. children born to women with T4 levels in the lowest 10th percentile of the normal range had a higher risk of low IQ and attention deficits. (Reviewed by Zoeller et al. 2002).

Most developmental neurotoxicity studies have focussed on general impairment of behaviour, but some studies have also found evidence for effects on sexual dimorphic behaviour. Especially, for EDCs with oestrogenic or antiandrogenic effects gender-related effects seems of relevance. Therefore, effects on one gender but not the other should not be dismissed, but be evaluated in the context of effects on sexual differentiation of the brain. In general, developmental neurotoxicity effects in animals are considered to indicate a potential for effects in humans.

2.5.4 Discussion

The OECD test guidelines for reproductive toxicity studies provide adverse effect data from endocrine and other mechanisms and are generally providing the most conclusive types of information for risk assessment. Generally, effects in these studies directly indicate potential for effects in humans.

The effects observed in reproductive toxicity studies may be due to other mechanisms than endocrine effects, however, the pattern of effects may indicate that endocrine effects are involved. For example, a pattern of decreased anogenital distance, retained nipples and effects on reproductive organs in males indicate that anti-androgenic effects are involved, while early sexual maturation in females in the absence of effects on body weight indicate that oestrogenic effects may be involved.

Reproductive toxicology comprises development toxicity and effects on male and female fertility. Developmental toxicity includes any effects interfering with normal development both before and after birth. These effects may arise due to exposure of the parent prior to conception, during prenatal exposure, and during postnatal development to the time of sexual maturation. The effects may be manifested during any time of the life span of the individual. Effects on male and female fertility include adverse effects on libido, sexual behaviour, any aspects of spermatogenesis or oogenesis, or on hormonal activity or physiological response which interferes with the capacity to fertilise, fertilisation itself or the development of the fertilised ovum up to and including implantation.

The vulnerable windows for developmental toxicity effects are prior to conception, during prenatal development and during postnatal development to the time of sexual maturation. As the development and maturation of the reproductive organs and functions takes place both pre- and postnatally it is important for the assessment of effects of EDC that the dosing period covers these periods.

The prenatal developmental toxicity study (TG 414) only covers effects induced during prenatal development and visible at birth and consequently this study can only reveal a

limited number of ED effects. In the one-generations study and the developmental neurotoxicity study, the exposure of the offspring is stopped at weaning and consequently juvenile animals are not exposed. The brain still undergoes development during the juvenile period, but effects induced during this period are not covered in the guidelines. The exposure period in the reproduction/developmental toxicity-screening test is from a few weeks before mating, during the pregnancy period, and until a few days after birth, i.e. 3-6 days. The two-generation study includes exposure of both sexes during all stages of reproduction, and in that sense this study is unique and the only study where the exposure period covers all vulnerable windows for effects of EDCs.

Developmental toxicity effects may become manifest at any time point in the life span of the organism. The regulatory test methods include some assessment of late effects especially in relation to reproductive function and developmental neurotoxicity.

The sensitivity of the prenatal developmental toxicity study (TG 414) for detection of rare events such as malformations is limited, due to the use of a relatively small number of animals. To assess the developmental toxicity of a chemical, it is therefore important to include information on other developmental effects such as minor anomalies, variations, foetal death and growth. In addition, malformations of organs developing after the period of major organogenesis, e.g. the reproductive organs and the brain, may not be detected if the study is performed according to the prior guideline for the teratology study. An example is the suspected endocrine disrupter dibutylphthalate where only exposure during the period of male sexual differentiation after gestation day 15 has resulted in major disturbances in the morphological and functional development of the male reproductive system (Mylchreest et al 1999). Consequently, in the updated OECD TG 414 the dosing period has been extended to cover at least the period from implantation to one prior to the day of scheduled kill, which is one day before the expected day of delivery.

The two-generation study requires that growth and survival of the offspring, sexual maturation, fertility, and semen quality and oestrus cyclicity are investigated in young adult animals. Malformations of reproductive organs such as hypospadias and cryptorchidism can be found in adolescent animals in the generation studies. However, only one offspring per sex per litter is selected at weaning and consequently the generation studies have a very limited sensitivity for detecting malformations of the reproductive organs. This is a potentially important limitation in the generation studies and therefore, it seems relevant to consider enhancement of the generation studies to include investigations of malformation in more than one offspring per sex per litter.

The behavioural functions assessed in the developmental neurotoxicity study cover many important aspects of the nervous system, however, some functions of relevance for endocrine disruption such as social interaction and mating behaviour are not included in guidelines at present.

Fertility effects are investigated by mating of animals to produce one litter per pair in the two-generation study, the one-generation study and the screening tests. Male rats, however, have a large excess of sperm compared to humans and therefore even major effects on sperm number may not result in decreased fertility. Consequently, the guideline for the two-generation study has recently been updated to include assessment of sperm quality. This assessment includes measures of sperm number, anomalies and motility, but the ability of the sperm to fertilise the ovum is not assessed. In order to increase the sensitivity for detection of fertility effects investigation of in vitro

fertilisation could be considered. Another potentially more sensitive possibility may be the continuous breeding protocol where the animals produce several litters instead of only one litter per pair.

Effects becoming manifest during ageing are not included in any guideline for reproductive toxicity. The reproductive span of females is limited from the time of sexual maturation to the onset of menopause. The time of sexual maturation may be assessed in the two-generation study, but menopause is not assessed. Consequently, effects on the reproductive span of females are not covered in reproductive toxicity guidelines.

Of critical concern is the possibility that developmental exposure may result in an acceleration of age-related decline in function. Animal studies have for example demonstrated that developmental exposure to neurotoxicants such as methyl mercury, methylazoxymethanol and ethanol may cause few or no neurotoxicity effects in young animals, but marked effects in ageing animals.

Another recently emerged concern is the so-called "Barker hypothesis" which is based on epidemiological studies showing that low birth weight or thinness at birth is associated with an increased risk of cardiovascular and metabolic disease as well as neuroendocrine dysfunction in adult life. These results support the notion that adverse effects in foetal life may permanently alter the structure and function of the adult offspring, a phenomenon dubbed 'foetal programming'. Indications may be observed in the two-generation study if the 'foetal programming' also results in effects on growth and survival of the animal, but there are no specific test methods investigating neuroendocrine dysfunction in adults.

In summary, among the OECD Test Guidelines for reproductive toxicity, all vulnerable periods of development are covered only in the two-generation study design. Late effects are partly covered in young adults, especially in relation to reproductive function and developmental neurotoxicity, but potentially important late effects are not assessed. Effects becoming manifest during ageing are not included in guidelines for reproductive toxicity.

Enhancement of the existing OECD TG for reproductive toxicity in order to increase the sensitivity for detection of ED effects are of high priority at present in the OECD. A number of potential enhancements for the detection of effects of EDCs seem relevant and lack of effects in reproductive toxicity studies can therefore at present not exclude the possibility for ED effects.

The following enhancements could be considered in order to increase the possibility and sensitivity for detection of EDC effects:

<u>TG 416, two-generation study:</u>
A number of studies have shown that EDCs can affect anogenital distance and nipple retention and consequently assessment of these endpoints in the offspring, i.e. in F1 and F2, should be considered.

Exposure to EDCs may cause malformations of the reproductive organs such as hypospadias and cryptochidism. These effects only become manifest during the postnatal development of the animals and will therefore not be detected in the prenatal developmental toxicity study. In order to increase the sensitivity for detection of such malformations to a similar level as malformations observable in the prenatal

developmental toxicity study, malformations in the reproductive organs should be investigated in more than one offspring per sex per litter in the two-generation study.

TG 414, prenatal developmental toxicity study:
A number of studies have shown that EDCs, especially some phthalates, can decrease the prenatal testosterone surge in males foetuses and that this effect may be involved in the effects observed postnatally on e.g. anogenital distance and reproductive organs. This effect on the testosterone surge can be measured a few days before birth (e.g. gestation day 21) in male foetuses and could without major changes in the design be included in the OECD 414. Therefore, it seems relevant to consider enhancement of the OECD 414 to include measurements of testicular testosterone levels in GD 21 foetuses.

TG 415, one-generation study:
In contrast to the two-generation study, the one-generation study has not been updated. In relation to EDC effects, it seems relevant to update the one-generation study to include similar endpoints as the two-generation study, e.g. sperm analysis, oestrus cyclicity and histopathology in paternal animals.

In addition, it would be relevant to consider extending the exposure period to at least sexual maturation instead of postnatal day 21 and include assessment of anogenital distance, nipple retention, sperm and oestrus cyclicity endpoints as well as histopathological investigations of offspring.

TG 426, developmental neurotoxicity study:
Hormones play a central role during development and sexual differentiation of the brain, and studies of EDCs have shown gender-related effects on behaviour. Consequently, in studies of potential EDCs assessment of effects on sexual dimorphic behaviour should have a high priority.

Some behavioural endpoints of relevance for EDC effects, such as mating and nursing behaviour, are not included at present and consequently it is recommended to consider enhancing the guideline to include these endpoints.

2.5.5 Conclusions at level 5

The reproductive toxicity studies provide adverse effect data and are especially useful for risk assessment as they indicate potential for effects in humans. The effects observed in reproductive toxicity studies may be due to other mechanism that endocrine effects, but the pattern of effects, e.g. decreased anogenital distance and malformations of reproductive organs in males, may indicate that endocrine effects are involved.

Among the OECD Test Guidelines for reproductive toxicity, exposure during all vulnerable periods of development is only performed in the two-generation study design. Late effects becoming manifest after weaning of the animals are partly covered in young adults, especially in relation to reproductive function and developmental neurotoxicity, but potentially important late effects are not assessed. Effects becoming manifest during ageing are not included in any guidelines for reproductive toxicity.

A number of enhancements of the OECD Test Guidelines for reproductive toxicity for the detection of effects of EDCs seem relevant and lack of effects in reproductive toxicity studies can therefore at present not fully exclude the possibility for ED effects caused by chemicals tested negative.

3 Specific chemicals and observed ED-effects

In the following, the LOAELs from assays at level 2-5 are compared using examples of some chemicals representing various modes of ED-action. It should be emphasised that, in order to get enough information on the various compounds, the studies listed in the table do not necessarily represent the dosing regime recommended by the guidelines. Data concerning the endocrine disrupting effect of nonylphenol, vinclozolin, linuron, procymidone, finasteride, some phthalates, and others are shown in Table 9. The references used in this chapter mainly originate from the published literature. No attempt have been made to include all references for each chemical, however, we have attempted to include the most relevant results.

The table also provides information on the current hazard classification for reproductive toxicity in EU and when available the NOAEL/LOAEL used in the EU risk assessment of the chemical.

Table 9. Comparison of data from various test methods mainly originating from published reports.

Chemical	**Test, LOAEL (in bold), route of administration**	**Effects**	**References**
Oestrogens			
Bisphenol A	*In vitro*: Oestrogen-related screening assays	Moderate oestrogen.	Andersen et al., 1999
	Uterotrophic: **100-300** mg/kg/day s.c.;375-600 mg/kg/day p.o.	Uterine wt↑	Kannon et al., 2003
	Hershberger: >1000 mg/kg/day p.o.	No effect.	Kim et al., 2002
	Pubertal male: **50** mg/kg/day s.c. PND 22-32	Prl ↑ lateral prostate wt↑ prostatitis.	Stoker et al., 1999 cited in Stoker et al., 2000
	Pubertal female: 50-400 mg/kg/day p.o. at **100**; 400-800 mg/kg/day p.o. At **600** and 800	Disturbed oestrus cycle, No effect on VO. Effect on VO.	Laws et al., 2000; Ashby and Tinwell, 1998
	TG 407: 600 mg/kg/day p.o.	Dioestrous prolonged, Alpha2uglobulin ↓	Yamasaki et al., 2002b
	TG 416: 0.015-7500 ppm in diet At 7500 ppm corresponding to **500** mg/kg/day	Delayed puberty in males and females and concomitantly decreased body weight.	Tyl et al., 2002
	EU-Risk assessment: NOAEL 50 mg/kg/day	Developmental effects.	EU-RAR, draft July 2002 NOAEL is based on results from Tyl et al., 2002
	Classification*	R62 Cat. 3	

4-nonylphenol	*In vitro*: Oestrogen-related screening assays	Moderate oestrogen.	Soto et al., 1995
	Uterotrophic: **15-35** mg/kg/day s.c.; 75 mg/kg/day p.o.	Uterine wt↑	Kannon et al., 2003
	Pubertal female: **50** –200 mg/kg/day p.o.	Advance VO, Disturbed oestrus cycle.	Laws et al., 2000
	TG 407: **200** mg/kg/day	Sperm count ↓	OECD-validation, phase two unpublished data
	TG 416: **50** –160 mg/kg/day p.o.	Oestrus cycle length, advanced VO, ovarian wt sperm/spermatid count.	NTP, 1997 cited in EU-RAR draft Sep. 1999
	Multigeneration study: 200-2000 ppm in diet 650 ppm corresponding to **30-100** mg/kg/day	Advance VO, oestrous cycle disruption, ovary weight ↓, epididymal sperm density in F2 males ↓	Chapin et al., 1999
	EU-Risk assessment: NOAEL 15 mg/kg/day		EU-RAR, draft Sep. 1999
	Classification	R62/63 Cat. 3	
4-*tert*-octylphenol	*In vitro*: Oestrogen-related screening assays	Moderate oestrogen.	Soto et al., 1995
	In vitro (testis testosterone secretion): 10-500 mg/l. At **100** mg/l:	T secretion↑	Haavisto et al., 2003
	Uterotrophic: 200 s.c.	Uterine wt↑	Yamasaki et al., 2002b
	Pubertal female: >**100** mg/kg/day p.o. Pubertal female: **200** mg/kg/day p.o.	Advance VO. Advance VO, disturbed oestrus cycle.	Gray and Ostby, 1998; Laws et al., 2000
	TG 416: 0.2 to 2000 ppm in diet. Reproductive toxicity: NOAEL 2000 ppm corresponding to **369-111** mg/kg/day	Body weight related delays in VO and PPS.	Tyl et al., 1999
	Classification	–	
AR antagonists			
Vinclozolin	*In vitro*: AR related screening assays	The metabolites are potent antiandrogens.	Kelce et al., 1994
	Hershberger: **10**-200 mg/kg/day p.o	Male reproductive accessory organ weights ↓	Nellemann et al., 2003
	Pubertal male:**10,** 30, 100 p.o. day 22-54 30 and 100	Serum LH↑ T↑ delayed PPS by 4-7 days.	Monosson et al., 1999
	Intact male: 10 -300 mg/kg/day p.o. **150** mg/kg/day	At 10: T_4 ↓ At 75: FSH, LH, PRL↑ At 150: reproductive organ wt ↓	O'Connor et al., 2002b
	Developmental: **12** mg/kg/day p.o.	AGD↓ nipple retention.	Hellwig et al, 2000
	Developmental **3,125**-100 mg/kg/day p.o. GD14-PND3	AGD↓ nipple retention, prostate wt↓, hypospadias.	Gray et al., 1999a
	Classification	R40, R60/61 Cat. 2	

Procymidone			
	In vitro (AR reporter gene assay):	Moderate antiandrogen.	Nellemann et al., 2003
	Hershberger: **10** mg/kg/day p.o.	Male reproductive accessory organ weights ↓	Nellemann et al., 2003
	Developmental: **25**-200 mg/kg/day p.o. GD14-PND3.	AGD↓ nipple retension, hypospadia cleft phallus, undescended testes, reduced weights of androgen dependent tissue.	Ostby et al., 1999
	Classification	[R40,R62/63] Cat. 3 or Cat. 2 or not classified.	
Prochloraz	*In vitro*:	Weak antiandrogen. Weak antioestrogen. Weak aromatase inhibitor.	Vinggard et al., 2002 Andersen et al., 2002 Andersen et al., 2002
	Hershberger: **50**-200 mg/kg/day p.o.	Accessory reproductive glands↓	Vinggaard et al., 2002
	Developmental **30** mg/kg/day p.o. GD7-PND17	T levels↓, nipple retention.	Vinggard et al., 2003 abstract
	Classification	-	
Linuron	*In vitro* androgen related screening assays:	Weak antiandrogen.	Lambright et al., 2000
	Hershberger: **10-30** mg/kg/day p.o.; Hershberger: **100** mg/kg/day p.o.	Reproductive accessory tissue weights ↓ Male reproductive accessory tissue weights ↓	OECD-validation phase two unpubl. data; Lambright et al., 2000
	Pubertal male: **40** mg/kg/day p.o.	Delayes PPS by 2.5 days	Gray et al., 1999b
	Intact male: 25–**150** mg/kg/day p.o. for 15 days	> 25: T_4↓ > 50: T_3↓ and E2↓ >100: DHT↓ At 150: Epididymides, accessory reproductive glands↓ testes histopat., T↓, PRL↓, LH↓	O'Connor et al., 2002b
	Developmental **12.5**-50 mg/kg/day p.o. GD 12-21	Epididymal abnormalities⇒ accumulation of tubular fluid and large testes. At 50 mg: permanent nipple retention and AGD↓	McIntyre et al., 2002
	Classification	R40 R61 Cat. 2, R62 Cat. 3	

Steroid synt. inhibitor			
DEHP	*In vitro*:	Weak oestrogen.	Soto et al., 1995
	Foetal steroidogenesis, **300**-750 mg/kg/day p.o. GD 7-21	T levels in serum and testes ↓, T production↓	Borch et al., 2003a,b
	Pubertal male: **2000** mg/kg/day p.o. PND 42-56	T↓ testes, vp, sv wgt↓	Oishi, 1995 cited in Stoker et al., 2000
	Young males: 5-5000 ppm in diet for 13 weeks 500 ppm equal to **37** mg/kg/day	Testes histopathology.	Poon et al., 1997
	Developmental 32.5 and 325µl/litre in drinking water corresponding to **3.0**-35 mg/kg/day GD1-PND21	Testes histopathology.	Arcadi et al., 1998
	TG 416: 1.5 (Control 1 and 2), 10-10000 ppm in diet. At 300 ppm corresponding to **14** mg/kg day in the F1 and F2 generation	Aplastic/small testes and epididymides.	Wolfe et al., 2003 unpublished, cited in EU-RAR draft August 2003
	EU-Risk assessment NOAEL 5 mg/kg	Reproductive effects.	EU-RAR draft August 2003
	Classification	R60/61 Cat. 2	
DBP	*In vitro*:	Weak oestrogen.	Soto et al., 1995 Harris et al., 1997
	Hershberger: **500**-1000 mg/kg/day p.o.	All male reproductive accessory organ weights ↓	Ashby and Lefevre 2000
	Uterotrophic: >1000 mg/kg/day p.o.	No effect on uterus weight.	Yamasaki et al. 2002b
	Pubertal male: **500** and 1000 mg/kg/day p.o. PND 21-57 Pubertal male: **500**	Serum LH and FSH↑, Testes, epididymides, sv, vp ↓/-	Gray et al., 1988 cited in Stoker et al., 2000 Ashby and Lefevre2000a
	Intact male rat: 250-**1000** mg/kg/day p.o. PND 70-85	>250: E2↓ and T_3↓, microscopic changes at 1000.	O'Connor et al., 2002b
	Developmental:100-500 mg/kg/day p.o. GD12-21 **100** mg/kg/day	Delayed PPS, AGD↓, nipple retention.	Mylchreest et al., 1999
	Developmental: 0.5-500 mg/kg/day p.o. GD12-21 **100** mg/kg/day	Critical effect: nipple retention.	Mylchreest et al., 2000
	EU-Risk assessment NOAEL 50 mg/kg/day	Reproductive effects.	EU-RAR, draft June 2001, based on Mylchreest et al., 2000
	Classification	R61 Cat. 2 R62 Cat. 3	
DINP	*In vitro*: Oestrogen-related screening assay	Weak oestrogen.	Harris et al., 1997
	Foetal steroidogenesis **300** mg/kg/day p.o. GD7-GD21	T↓, testes histopathology.	Borch et al., 2003a, b
	Developmental 300-750 mg/kg/day p.o. GD7-PND17. **600** mg/kg/day	AGD↓, nipple retention.	Hass et al., 2003
	EU-Risk assessment LOAEL 114 mg/kg/day	Decreased maternal body weight and hepatic changes.	EU-RAR draft May 2001
	Classification	–	

Aromatase inhibitor			
Fenarimol	*In vitro* (cyp19 aromatase activity):	Weak to moderate aromatase inhibitor. Weak antiandrogen. Weak oestrogen.	Andersen et al., 2002 Andersen et al., 2002 Vinggaard et al., 1999
	Developmental:0 and 350 ppm in diet corresponding to 20 mg/kg/day, GD 0- GD 21 or until PND 3.5 or 5.	Inhibit oestrogen biosynthesis in the brain.	Hirsch et al., 1987
	Classification	R62/63/64 Cat. 3	
Fadrozole	*In vitro*: (cyp19 aromatase activity)	Potent aromatase inh.	Wade et al., 1994
	Pubertal female: >**0.6**-6.0 mg/kg/day p.o. PND 21-40	Delayed VO, uterine wts↓	Marty et al., 1999
	Pubertal Male: 0.6 and 6.0 mg/kg/day **p.o.?**	No effect.	Marty et al., 2001
	Intact male 1-25 mg/kg/day p.o. **5** mg/kg/day	At 1: T, DHT, T_4 ↓, at 5: Wt of androgen dependent tissue↓, at 25: FSH↑,	O'Connor et al., 2002c
	Classification	–	
5-α reductase inhibitor			
Finasteride	Hershberger: **0.2** mg/kg/day p.o.	All male reproductive accessory organ weights ↓	OECD-validation, phase two unpublished data;
	Intact male: dose **20** and 80 mg/kg /day p.o.	Accessory reproductive gland ↓	Marty et al., 2001
	Pubertal male: **25** mg/kg/day p.o.	Ventral prostate, seminal vesicles, and Cowper's gl. weight ↓	Ashby and Lefevre, 2000
	Developmental: 0.0003 to 300 mg/kg/day p.o. GD6-GD20 2.1 **%** decrease in AGD at **0.0003** mg/kg/day	AGD↓ nipple retention, hypospadia cleft prepuce.	Clark et al., 1990a
	Developmental: 0.0003 to 3 mg/kg/day p.o. GD15-PND21 13 **%** decrease in AGD at **0.03** mg/kg/day	AGD↓ nipple retention, hypospadia cleft prepuce, delayed preputial separation.	Clark et al., 1993
	Developmental: 0.01 to 100 mg/kg/day p.o. GD12-GD21 8 **%** decrease in AGD at **0.01** mg/kg/day	AGD↓ nipple retention, hypospadia cleft prepuce, prostate agenesis.	Bowman et al., 2003

Thyroid modulator			
PBDE (DE-47)	Females, 7 weeks old: 1, **6** and 18 mg/kg/day p.o. for 14 days Light microscopy of thyroid gland *ex vivo* measurement of microsomal enzyme activity (metabolic activation-EROD). *ex vivo* binding of 125I-T4 to plasma proteins (periferal TH transport). (LOAEL based on EROD measurements, TTR measure: LOAEL was 18 mg/kg/day)	No degenerative alterations. At 6 mg/kg/day: Induction of microsomal enzyme activity. Decrease in *ex vivo* binding of 125I-T_4 Activation of thyroid epithelia.	Hallgren and Darnerud, 2002
	Female SD-rats 7 weeks old: **18** mg/kg/day (DE-47) p.o. for 14 days (18 was the lowest dose used)	Decreased plasma free and total thyroxine (T_4) levels. By contrast, thyroid-stimulating hormone (TSH) levels were not significantly changed in any of the groups.	Hallgren et al., 2001
PBDE (DE-71)	Dams LE-rats: 1, **10** and 30 mg/kg/day p.o. from GD 6 to PND 21	No effect on maternal body weight gain, litter size or sex ratio, no effects in any measures of offspring viability or growth. Foetal/pup Serum T_4↓ (GD20-PND14) no effect on PND 36, Maternal T_4 ↓ (GD20-PND22), Foetal/pup T_3 no effect Decreased thyroxine levels.	Zhou et al., 2002
	Mice: 0.8 (pure PBDE/ congener BDE 99) p.o. to pups on day 10 p.n. examined at 2-4 months of age.	Impaired habituation and learning and memory functions.	Eriksson et al., 2001

*The criteria for classification and the risk-sentences are described in table 12

[a] The evaluation of mechanisms and potency in the *in vitro* studies is based on a comparison with other environmental relevant chemicals and not a potent positive control chemical.

VO	vaginal opening
HP	hypothalamus
E	oestradiol
PPS	preputial separation
p.o.	oral dosing by gavage
SV	seminal vesicles
VP	ventral prostate
PND	post natal day
↓	decrease
↑	increase
↓/-, ↑/-	Variable results
EU-RAR	European Union Risk assessment Report (Draft)
Wgt	weight

Developmental: Level 5 studies, where time of exposure includes the foetal and/or lactational period.

[ ]ongoing evaluation of classification

3.1 Comparison of LO(A)ELs in EDC assays

In the following, the LOAELs in EDC assays are compared where data are available from in vivo screening assays as well as developmental studies. Consequently, the aromatase inhibitors fenarimol and fadrozole and thyroid modulators are not included.

Bisphenol A
As seen in table 9, the LOAEL observed in the pubertal male assay after s.c. dosing is 50 mg/kg. This LOAEL is lower compared to the tests, where oral dosing was used, i.e. the LOAELs are 100-600 mg/kg. This may be explained by the high first by pass metabolism of BHA in the liver.

Concerning EU risk assessment, the study by Tyl et al., 2002 has provided data on which the critical NOAEL of 50 mg/kg/day is based. Vom Saal (EU-RAR) has found reproductive effects of BHA at much lower doses. In scientific and in regulatory circles, these results as well as result obtained in order to reproduce the studies by vom Saal have been extensively debated. At present no conclusions can be drawn, but it is agreed that further studies are needed. Thus, at present it is not possible clearly to evaluate whether the NOAEL in the risk assessment of BHA will protect against endocrine disruption.

Nonylphenol
Except from the TG 407, where the LOAEL is 200 mg/kg, the LOAELs in the in vivo screening assays as well as in the multi-generation studies are within the same order of magnitude, i.e. 15-50 mg/kg. The Uterotrophic assay has a LOAEL around 15 mg/kg/day, while the NOAEL used in the EU risk assessment is 15 mg/kg/day based on a two-generation study. This is the only compound where the uterotrophic assay actually seems to be more sensitive than the two-generation test.

Octylphenol
All *in vivo* screening assays and the two-generation study have LOAELs within the same order of magnitude, i.e. around 100-200 mg/kg.

Vinclozolin
The Hershberger and pubertal male assays have similar LOAELs of 10 mg/kg. However, the intact male assay has a LOAEL of 150 mg/kg, which is 15 times higher. The LOAELs for developmental toxicity is around 3-12 mg/kg, i.e. 1-3 times lower compared to the Hershberger and pubertal male assays, and 13-50 times lower compared to the intact male assay. Hence, in case data from developmental studies were missing, it would have been sufficient to add an additional safety factor of e.g. 3, when using data from the Hershberger and pubertal male assays. However, a much larger safety factor around 50 is needed if data are only available from the intact male assay.

Procymidone
The LOAEL of 10 mg/kg obtained in the Hershberger test is 2.5 times lower compared to the LOAEL of 25mg/kg/day in the developmental study. However, the 25 mg/kg was the lowest dose tested for developmental toxicity. Hence, there is reason to believe that the LOAELs, which may be obtained by using these two study-designs, would be comparable.

Procloraz
In the Hershberger assay as well as in the developmental study, effects were found at dose levels of 50 and 30 mg/kg, respectively. In these studies, NOAELs were not found. Nevertheless, the dose levels causing effects in both designs seem quite comparable.

Linuron
The LOAELs in the Hershberger and the pubertal male assay are 10 and 40 mg/kg, respectively. In the intact male assay, a LOAEL of 150 mg/kg was observed. The LOAEL in the developmental study is 12.5 mg/kg, i.e. of similar magnitude as the LOAELs in the Hershberger and the pubertal male assay. However, the LOAEL obtained in the intact male assay is one order of magnitude higher compared to the other studies. In this case, an additional safety factor of 12 would protect against lack of developmental data needed for risk assessment.

Phthalates
For DEHP, the LOAEL, from the pubertal male assay is 2000 mg/kg, while the other assays at level 3 and 4 provide LOAELs ranging from 37-300 mg/kg. The developmental studies provide LOAELs ranging from 3-14 mg/kg, i.e. at least 100 times lower than the dose causing effect in the pubertal male assay and 10-20 times lower compared to the other assays at level 3 and 4. In the EU risk assessment of DEHP, an unpublished study by Wolfe et al. (unpublished cited in EU-RAR draft, 2003) found effects within the same low dosing area as earlier reported. Thus, concerning DEHP, the risk assessment of DEHP should with the data available protect against endocrine disruption.

For DBP, the LOAELs in the Hershberger and the pubertal male assay are similar, i.e 500 mg/kg. In the intact male assay, the LOAEL is 1000 mg/kg, i.e. twice as high as the other screening assays. In the developmental studies, the LOAELs are 100 mg/kg in both studies. Thus, the LOAELs from developmental studies are two times lower compared to the Hershberger and the pubertal male assay and 10 times lower compared to the intact male assay.

Earlier published in vitro data have demonstrated that the phthalates have very weak oestrogenic activity, while recent investigations has revealed effects of DEHP, DBP and DINP on foetal steroidogenesis (Parks et al., 2000; Mylchreest et al., 2002; Borch et al., 2003a, b).

Concerning DINP, the LOAEL of 114 mg/kg in the EU-RAR is based on parental effects rather than reproductive and developmental effects. However, Borch et al. (2003a, b) did not find a NOAEL in their study showing decreased testosterone synthesis in male foetuses at 300 mg/kg, and consequently effects on testosterone production may also be found at lower doses.

Finasteride
The LOAEL in the Hershberger assay is 0.2 mg/kg while the LOAELs in the intact male and pubertal male assay are 100 times higher i.e. 20 and 25 mg/kg, respectively. In developmental studies, the LOAELs range from 0.0003-0.03 mg/kg. Consequently, finasteride is another example of a very large difference between the LOAELs observed in the in vivo assays at level 3 and 4 compared to the level 5 studies, as the LOAELs are 100 to 1000 times lower in the developmental studies. This means that a risk assessment based on level 3 and 4 assays would most probably not provide sufficient protection against endocrine disruption and even an additional safety factor of 10 would be

insufficient in a situation, where no information about developmental toxicity effects is available.

3.2 Discussion and conclusions

Generally, looking across the data, the in vitro assays seem quite useful for supplementing the interpretation of in vivo studies with additional mechanistic information.

The intact male assay and the TG 407 seem to have a lower sensitivity compared to the other *in vivo* assays at level 3 and 4, and this is important to keep in mind when studies are evaluated. TG 407 is designed to detect any potential for toxic effects and therefore the doses used are often quite high. This together with the low number of males and females per group may partly explain the lower sensitivity compared to the other in vivo assays.

The Uterotrophic, the Hershberger assay as well as the pubertal male and female assays in most cases detect effects within the same range of dose levels as the level 5 tests. Thus, effect levels found in the mentioned assays are, at least for these endocrine disrupters, able to provide a relatively good prediction of the LOAELs for developmental toxicity. Therefore, we suggest that, in case the data is limited to investigations in the above-mentioned screening tests, including the Uterotrophic and the Hershberger assay, these assays can provide dose information useful for preliminary risk assessment of chemicals. Consequently, these data could be considered for regulatory considerations until data from further studies are available.

Generally, compared to fully developed animals, foetuses and young individuals may be more vulnerable to endocrine disrupters and the effects may be irreversible. An additional safety factor of e.g. 10 would in most cases be appropriate to compensate for lack of information from level 5 studies. However, there are some exceptions such as the 5-alpha reductase inhibitor finasteride, and the steroid synthesis inhibitor DEHP, where an additional safety factor needs to be much higher than 10 to provide sufficient protection until further results are obtained in the reproductive and developmental toxicity studies. Nevertheless, in case of lack of information, it should still be considered to use the LO(A)ELs obtained in the in vivo tests at level 3 and 4 for making a preliminary risk assessment including an additional safety factor, until data from further testing become available.

It should be emphasised that there are strain and species differences in laboratory animals (Hossaini et al., 2003; Spearow et al., 1999) and this issue should always be included in evaluation of a chemical.

4 Discussion and conclusions

4.1 Summary on the assays at level 2 to 5 and regulatory proposals

The following considers the information on mechanism, adverse effects and potency that may generally be obtained from the assays at levels 2-5 (see chapter 2 for details) combined with the results obtained for a number of studies on the effects of EDCs in these assays (see chapter 3 for details). In addition, possibilities for regulatory actions are proposed.

Examples of legislative instruments considered are inclusion on the EU-list of potential endocrine disrupting chemicals prioritised for further testing, prioritisation for further testing, hazard classification, risk reduction based on concern identified via risk assessment and/or inclusion of the substance in the authorisation procedure of REACH. A hazard classification of a chemical as toxic to reproduction in Cat. 1 or 2 leads in the existing EU legislation immediately to down stream consequences in other EU directives on worker protection, cosmetics, air quality, waste, export etc. According to the directives such chemicals should not be used by consumers, workers, pregnant and breast feeding workers or young workers.

Tables 10 and 11 give an overview of proposed regulatory actions in cases where endocrine disrupting effects have been found or not found, respectively.

It is important to keep in mind that evaluation of EDC data requires expertise, because of the complexity and multitude of ED effects and the rapid expansion of our knowledge on EDCs. Therefore, effects of potential EDCs should be evaluated by experts in the field on a case-by-case basis.

Level 2, *in vitro* assays providing mechanistic data

Positive *in vitro* test results from well-performed studies indicate potential EDC activity *in vivo*. The mechanism of action, e.g. oestrogenic or antiandrogenic activity, is generally considered relevant for humans. Therefore, reliable *in vitro* data can be used to place the chemical as a category 2 substance on the EU list of potential EDCs prioritised for further testing.

Results from *in vitro* tests may also be useful when considering hazard classification of a chemical. For example, if reproductive toxicity studies have demonstrated some effects, but the data is evaluated as only sufficient to place the chemicals in reproductive toxicity category 3, results from in vitro tests may be used as supportive evidence for upgrading to category 2.

In vitro data can provide valuable mechanistic data that is useful for the design of further *in vivo* studies. The potency of *chemicals in vitro* is generally difficult to extrapolate to the *in vivo* situation, but chemicals with high potency *in vitro* may also be expected to have a high potency *in vivo*.

It is debated whether results of *in vitro* tests at present *are* sufficient for prioritising chemicals for further testing *in vivo*, because a negative result does not exclude the possibility for EDC activity through other mechanisms. We find that if positive *in vitro*

data exist and humans are exposed to a substantial degree, further *in vivo* testing should be given a high priority. Regulatory actions until relevant *in vivo* results become available might also be considered if the potency of a chemical *in vitro* is of similar magnitude as well known EDCs.

Negative *in vitro* test results cannot be used to exclude potential ED activity because of limitations such as inability or unknown capacity to metabolically activate/deactivate chemicals and because ED activity can occur through mechanism other than those tested in the *in vitro* test systems.

QSAR models for ED activity and reproductive toxicity effects are under development. When reliable QSAR models have been developed, they may be used for prioritisation exercises as well as they may support experimental findings.

Table 10. Proposed regulatory actions after positive test outcome in tests assessing ED activity and reproductive and developmental toxicity

Type of assay	**Mechanism**	**Adverse effects**	**Potency**	**Regulatory action(s)**
Level 2 *In vitro*	Endocrine disruption, which is considered relevant for humans.	Uncertain, however a number of examples indicate possibility.	Difficult to extrapolate to *in vivo*, however, high potency *in vitro* may indicate high potency *in vivo.*	- Candidate for *in vivo* testing. - List of potential EDCs. - May support upgrading of classification from Rep. cat. 3 to Rep. cat. 2. - If high potency, the need for risk reduction may be considered.
Level 3 Short term *in vivo*	Endocrine disruption, which is considered relevant for humans.	Indication of adverse effects in reproductive toxicity studies.	Provide NOEL/LOEL for the endpoints investigated. LOELs seems to be of similar magnitude or higher than LOAELs in reproductive toxicity studies.	- High priority for reproductive and developmental toxicity testing. - List of potential EDCs. - Consider need for risk reduction until results of further studies are available (may be for several years). - Consider classification in (at least) cat. 3 for reproductive toxicity. - May support upgrading of classification from Rep. cat. 3 to Rep. cat. 2
Level 4 *In vivo* assays	ED, if type or pattern of effects observed.	Adverse effects or effects via a mechanism relevant for humans.	Provide NO(A)EL and LO(A)ELs.	- List of potential EDCs if supported by types of effects, *in vitro* or short term *in vivo* results. - Classification: upgrade from cat. 3 to cat. 2 if mechanism is ED. - Consider need for risk reduction.
Level 5 Reproductive and developmental toxicity studies	ED, if type or pattern of effects observed.	Provide evidence of adverse effects.	Provide NOAEL and LOAELs for reproductive and developmental toxicity effects.	- List of EDCs if supported by types of effects, *in vitro* or short term *in vivo* results. - Classification: upgrade from cat. 3 to cat. 2 if mechanism is ED. - Consider need for risk reduction.

Table 11. Proposals for regulatory actions after negative test outcome in tests assessing ED activity and reproductive and developmental toxicity

Type of assay	Evaluation	Regulatory action(s)
Level 2 *In vitro*	Cannot be excluded as potential EDC as much mechanism can be involved in ED effects.	Need for *in vivo* testing to be considered based on human exposure levels.
Level 3 Short term *in vivo*	Cannot be excluded as potential EDC as much mechanism can be involved in ED effects.	Need for reproductive and developmental toxicity testing to be considered based on human exposure levels.
Level 4 *In vivo* assays	Cannot be excluded, as potential EDC as the testing do not include exposure during e.g. development.	Need for reproductive and developmental toxicity testing to be considered based on human exposure levels.
Level 5 Reproductive and developmental toxicity studies	Probably not EDC if studies have included all relevant ED sensitive *in vivo* endpoints. If positive results from *in vitro* or other *in vivo* assays: Evaluated on a case-by-case basis.	Consider need for risk reduction in order to have a sufficient margin of safety. Consider need for further testing, e.g. toxicokinetics.

Level 3, *in vivo* assays providing data on single endocrine mechanisms

The Uterotrophic assay appears to be reliable for identifying substances with both strong and weak (anti)oestrogenic mode of action. The sexually immature female assay and the ovariectomized mature female assay have shown equivalent results in multi-laboratory comparison study within OECD. Clear evidence supports that an increase in uterine weight is caused by an oestrogen mode of action. Likewise, the Hershberger assay, even though it is still under validation within OECD, seems to be reliably in identifying substances with an (anti)androgen mode of action. The methods are still under validation within the OECD program, and the uterotrophic assay is in the final stage. The mechanisms involved in both assays are highly relevant for humans. Also, the effects on most of the target tissues in both assays are relevant for humans. A positive result, in for example the Uterotrophic assay indicate that the substance in a developmental and reproductive study would cause an early vaginal opening in female pups, i.e. adverse effect on puberty onset. Therefore, results from these studies can be used to place the chemical as a category 1 substance on the EU list of potential endocrine disrupting chemicals.

It has been argued that effects found in the Uterotrophic and Hershberger assays demonstrate a mechanism and is therefore only useful as a tool for the prioritisation of chemicals for further testing. The present use of the two assays is mainly to demonstrate the agonistic or antagonistic characteristic of chemicals in in vivo in screening assays. A positive response in any of the two assays suggests the need for the substance to advance to reproductive and developmental toxicity testing for adverse effects. The knowledge concerning the activity of the chemical, e.g. oestrogenic or antiandrogenic, should be used when considering the design of further studies.

The Uterotrophic and Hershberger assay, however, do provide in vivo NOEL/LOELs for the endpoints examined. For several EDCs, the dose levels causing effects in these assays seem to be of a similar magnitude or higher than those causing effects in reproductive and developmental toxicity studies (see chapter 3). As results from reproductive and developmental toxicity testing may take several years to obtain, it seems warranted to perform preliminary risk assessment of the chemicals based on the

NOEL/LOELs from the Uterotrophic and Hershberger assays. The use of a larger margin of safety should be considered, because developmental toxicity effects may occur at lower dose levels than those causing effects in these short-term assays.

Results from the Uterotrophic and Hershberger assay may also be useful when considering hazard classification of a chemical. For example, if reproductive toxicity studies have demonstrated some effects that may be ascribed to an antiandrogenic mechanism, but the data is evaluated as only sufficient to place the chemicals in reproductive toxicity category 3, positive findings in the Hershberger assay may be used as supportive evidence for upgrading to category 2. In the absence of data from reproductive and developmental toxicity studies, data from the Uterotrophic and Hershberger assay may support classification for reproductive toxicity in category 3, because the results indicate a high probability for reproductive and developmental toxicity effects at similar or lower dose levels.

A negative uterotrophic response, in a thorough doses-response study, indicates that the test compound is not an ER-ligand. Equally a negative response in the Hershberger assay indicates that the test compound is neither an AR-ligand nor a 5-alpha reductase inhibitor. This means that a test compound found negative in for instance the Hershberger assay may still have endocrine disrupting properties mediated through another mechanism.

The ECT-test on steroidogenesis provides information on the ability of a substance to be an ED acting on the production of steroids, incl. sex hormones. Positive results in this test indicate that the substance can cause adverse effects in a pubertal and/or a two-generation test. The ex vivo methods are used to assess substances for altering steroid production and secretion (EPA 2002). Generally, these ex vivo methods are considered to be more useful than an in vitro assay, because the method combines *in vivo* exposure with an *in vitro* assessment of the effect. The limiting factor for the relevance of the test is the inclusion of the *in vitro* assessment of the effect, which does not take account of for example the HPA. A positive result in the *ex vivo* studies indicate a potential for effects *in vivo* and will as such give some basis for concern, however the potency of the chemical is difficult to determine. Generally, the results of these tests can be used to consider placing the chemical as a category 1 substance on the EU list of potential endocrine disrupting chemicals prioritised for further testing. In addition, results demonstrating clear effects on production of steroids, especially the testosterone surge during prenatal development of the males, may be used similarly as results from the uterotrophic or Hershberger assay for preliminary risk assessment and for hazard classification purposes.

At present, it is uncertain to what extent the frog metamorphosis assay can be used for screening for effects in humans due to the limitations of the method.

Level 4, *in vivo* assays providing data about multiple endocrine mechanisms and effects

The pubertal assays provide information about the potency of the compound in vivo. Effects on endpoints included in the assays can either be considered as adverse or represent an effect on a mechanism relevant for humans. Therefore these assays can be used to provide NO(A)ELs/LO(A)ELs to be used in human risk assessment and for classification. Also, positive findings can be used to place the chemical as a category 1 on the EU list of potential endocrine disrupting chemicals.

The sensitivity of these assays may be improved by making the optional hormone investigations mandatory in the protocol. This would aid in characterising the mode of action of the investigated endocrine disrupter. Such information will be useful for the design of reproductive and developmental toxicity studies.

The enhanced TG 407 provides a thorough assessment of the potential endocrine disruption effects of a chemical in young adult animals. Effects on most endpoints included in this assay can be considered as adverse and relevant for humans. Therefore results from this assay can be used for human risk assessment and for hazard classification.

Overall, the tests included at level four provide information about the potency of the compound in vivo. The sensitivity of the intact male assay seems to be somewhat lower compared to the other tests. However, inclusion of hormone investigations in this assay may increase the sensitivity. Effects on the various endpoints included in the assays can either be considered as adverse or represent an effect on a mechanism relevant for humans. Therefore, these assays can be used to provide NO(A)ELs/LO(A)ELs to be used in human risk assessment. The use of a larger margin of safety should be considered, because developmental toxicity effects in some cases have been shown to occur at lower dose levels than those causing effects in these assays.

Level 5, *in vivo* assays providing adverse effect data from endocrine and other mechanisms

The reproductive and developmental toxicity studies provide adverse effect data, which are used for risk assessment and classification, because the results indicate potential for harmful effects in humans.

The effects observed in reproductive toxicity studies may be due to other mechanism than endocrine disruption. The pattern of effects may however indicate that endocrine effects are involved. For example, a pattern of decreased anogenital distance, retained nipples and effects on reproductive organs in males indicate that antiandrogenic effects are involved, while early sexual maturation in females in the absence of effects on body weight indicate that oestrogenic effects are involved. In such cases, the results can support classification of the chemical in category 2 for reproductive toxicity and inclusion of the chemical as a category 1 substance on the EU list of potential endocrine disrupting chemicals.

A number of potential enhancements for the detection of effects of EDCs seem relevant and lack of effects in reproductive toxicity studies can therefore at present not exclude the possibility for ED effects. If no effects are found in an enhanced two-generation study, where all relevant ED sensitive endpoints are assessed, the chemical is most probably not an EDC. However, if there are positive results from in vitro studies or other in vivo assays, there is a need for further studies to clarify the discrepancy. These further studies may for example include studies of toxicokinetics and/or the use of other strains of animals.

4.2 Enhancement of existing OECD TG

Enhancement of the existing OECD TG for reproductive toxicity in order to increase the sensitivity for detection of ED effects are of high priority at present in the OECD. Based on considerations of sensitivity to EDC effects, a number of relevant enhancements were identified in chapter 2. The need for these enhancements should be considered in

relation to the testing strategy used. For example, enhancement of the one-generation study may not be needed, if the two-generation study is chosen as the test to be used for assessing potential EDCs for reproductive and developmental toxicity.

The proposed enhancements include:

- TG 407 repeated dose 28-day oral toxicity study in rodents (enhanced):
  - Spermatogenic staging in the histopathological examination of the testes, in order to compensate for not dosing throughout a complete spermatogenic cycle.
- TG 414, prenatal developmental toxicity study:
  - Consider including measurements of testicular testosterone levels in GD 21 foetuses.
- TG 415, one-generation study:
  - Update to include similar endpoints as the 2-generation study, e.g. sperm analysis, oestrus cyclicity and histopathology in paternal animals.
  - Consider extending the exposure period to postnatal day 90 instead of day 21 and include assessment of anogenital distance, nipple retention, sperm and oestrus cyclicity endpoints as well as histopathological investigations of offspring
- TG 416, two-generation study:
  - Assessment of anogenital distance and nipple retention in F1 and F2
  - Investigations of malformations of the reproductive organs in more than one offspring per sex per litter.
- TG 426, developmental neurotoxicity study:
  - Evaluate effects on sexual dimorphic behaviour
  - Include assessment of e.g. mating behaviour.

4.3 Hazard classification for EDC effects and authorisation

EU chemicals policy

The European Commission's proposal for new chemicals legislation states that chemicals meeting certain criteria for very high concern should be brought within a prior authorisation scheme. This could bring in the presumption against the use of such substances. The Commission proposed that substances classified as CMRs (carcinogens, mutagens or substances that are toxic to reproduction) should be subject to the authorisation scheme.

The existing criteria for hazard classification for reproductive and developmental toxicity in EU are shown in table 10, whereas the GSH criteria are shown in table 11.

The European Commission's White Paper "Strategy for a future Chemicals Policy" states that the majority of endocrine disrupting chemicals would have to undergo authorisation. It is highlighted that the health effects which have so far been associated with endocrine disrupting chemicals would qualify a substance either to be classified as a carcinogenic or as toxic for reproduction and so would trigger its submission to authorisation. Unfortunately, this may be an optimistic interpretation of the scope of the

current proposals. For example, some substances known to have endocrine disrupting properties in mammals would not be caught by the current proposal to bring CMRs into the authorisation process. One reason is that the Commission's proposal only relates to CMR categories 1 and 2 (see table 12 for an explanation of these categories for reproductive and developmental toxicity). This would for example exclude Bisphenol A and nonylphenol from being subjected to authorisation, because these chemicals are currently classified as toxic to reproduction in category 3.

The recently updated EU Technical Guidance Document for Risk Assessment (EU-TGD) states that:

"An endocrine disrupter is an exogenous substance that causes adverse health effects in an intact organism or its progeny through alterations in the function of the endocrine system. Thus, endocrine disruption is a mechanism rather than an adverse health effect.

The concern for endocrine disruption has resulted in the development of newly proposed test guidelines to specifically address effects on hormone homeostasis and on male and female reproductive organs. In addition, suggestions are being considered for new parameters to be incorporated into existing test guidelines. At the time when this TGD was being written several validation studies were still ongoing, precluding recommendations for the use of new test protocols. With respect to endocrine disruption, the two-generation study (OECD TG 416) is currently the most complete study available. Both in this study and in the developmental toxicity study (OECD TG 414), additional endocrine-sensitive parameters may be studied on a case-by-case basis when endocrine disruption is an issue of concern."

As such this wording in the EU TGD seems to reflect a similar approach as the White Paper, i.e. that at present hazard classification is to be based on adverse effects and not on ED mechanisms.

WWF's initial proposals for the regulation of chemicals with endocrine disruption properties

A different approach to the new EU chemicals legislation has been brought forward by WWF in a Discussion Paper in 2002. It is suggested by WWF that EDCs should be placed in sub-categories (similar to those currently used for carcinogens, mutagens or substances toxic to reproduction) depending on the level of available evidence for endocrine disruption.

The categories proposed are:

Category 1A EDC Substances that alter the function of the endocrine system and consequently cause adverse health effects in an intact organism

Category 1B EDC Substances as defined is 1A, but where the causal mechanism is still unclear, but strongly suspected to be mediated via disruption of the endocrine system

Category 2 EDC Substances where there is less evidence for endocrine mediated effects, and/or where endocrine disruption is strongly suspected or known but where there is debate whether the effects reported should be considered as adverse

Category 3 EDC Substances suspected to be endocrine disrupters on the basis of simple *in vitro* tests for endocrine disruption or non-validate QSAR screens, unless there are data sufficient to negate the concern.

WWF considers that category 1 and 2 EDCs should be brought within the authorisation system. Category 3 should also be able to be brought under the system if within set time frames, data are still lacking to negate the concern.

When comparing the WWF proposal to the EU criteria for classification for reproductive toxicity, it seems like most of the chemicals in category 1 and 2 EDC could be classified for example for reproductive toxicity in category 2. If the effects observed are cancer the same situation probably exists. Therefore, we find that a more feasible choice could be to update the use of the existing criteria for hazard classification to include more considerations of the data on EDC activity and effects.

Use of EDC data in existing criteria for reproductive toxicity classification

The adverse effects of some EDCs would qualify a substance either to be classified as a carcinogenic or as toxic for reproduction in category 1 or 2 and so would trigger its submission to authorisation. However, as already mentioned some EDCs and suspected EDCs only qualify for hazard classification in category 3 or no classification and are therefore not triggered for submission for authorisation.

Instead of developing a new set of classification criteria for EDCs as proposed by WWF, we find that an extended use of the available data on ED mediated effects when considering hazard classification could be more feasible and relevant. The background for this is that the effects of EDCs hopefully could be covered by the classification of CMRs, if the guidance for the use of these criteria were updated.

Concerning hazard classification for reproductive toxicity, it is therefore proposed:

- evidence of EDCs activity from *in vitro* testing or *in vivo* screening may be used as 'other relevant information' demonstrating e.g. that the chemicals operates with a mechanism relevant for humans. This information can support upgrading from category 3 to category 2, for example in cases where it is debated whether reproductive and developmental toxicity effects should be considered as adverse.
- evidence of EDC activity from *in vivo* screening models such as e.g. the Hershberger or Uterus test may be used directly for classification, because the chemicals are suspected to cause reproductive and developmental toxicity at dose levels of similar or lower magnitude. Consequently, it seems warranted to classify such chemicals for reproductive toxicity in at least category 3.

In the first situation, the upgrading from category 3 to 2 means that the chemical will be triggered for submission for authorisation. In the second situation, however, chemicals placed in category 3 based on *in vivo* evidence of ED activity and a strong suspicion of potential developmental toxicity will not automatically be triggered for authorisation. Based on the strong suspicion of developmental toxicity effects, we find that authorisation of these chemicals seems warranted until further test data that are sufficient for deciding whether the chemical should be classified in category 2 or not have been provided.

At present, the limitations concerning *in vitro* testing for EDC activity means that *in vitro* results alone are insufficient basis for classification. However, such chemicals should certainly be on the EU list of potential EDCs and prioritised for further testing, if the exposure of humans is not negligible.

Table 12. Existing EU criteria for classification and labelling of substances toxic to reproduction

Category 1:
Substances known to impair fertility in humans
There is sufficient evidence to establish a causal relationship between human exposure to the substance and impaired fertility.
50
Substances known to cause developmental toxicity in humans
There is sufficient evidence to establish a causal relationship between human exposure to the substance and subsequent developmental toxic effects in the progeny.

The following symbols and specific risk phrases apply for category 1:
For substance that impair fertility in humans: T; R60: May impair fertility
For substances that cause developmental toxicity: T; R61: May cause harm to the unborn child

Category 2
Substances, which should be regarded as if they impair fertility in humans
There is sufficient evidence to provide a strong presumption that human exposure to the substance may result in impaired fertility on the basis of:

- Clear evidence in animal studies of impaired fertility in the absence of toxic effects or evidence of impaired fertility occurring at around the same dose levels as other toxic effects but which is not a secondary non-specific consequence of the other toxic effects.
- Other relevant information.

Substances which should be regarded as if they cause developmental toxicity to humans
There is sufficient evidence to provide a strong presumption that human exposure to the substance may result in development toxicity, generally on the basis of:

- Clear results in appropriate animal studies where effects have been observed in the absence of signs of marked maternal toxicity, or at around the same dose levels as other toxic effects but which are not a secondary non-specific consequence of the other toxic effects.
- Other relevant information.

The following symbols and specific risk phrases apply for category 2:
For substances that should be regarded as if they impair fertility in humans: T ; R60 : May impair fertility
For substances that should be regarded as if they cause developmental toxicity in humans: T ; R61 : May cause harm to unborn child.

Category 3
Substance which cause concern for human fertility
Generally on the basis of:

- Results in appropriate animal which provide sufficient evidence to cause a strong suspicion of impaired fertility in the absence of toxic effects, or evidence of impaired fertility occurring at around the same dose levels as other toxic effects, but where the evidence is insufficient to place the substance in Category 2.
- Other relevant information.

Substances which cause concern for human owing to possible developmental toxic effects
Generally on the basis of:

- Results in appropriate animal studies which provide sufficient evidence to cause a strong suspicion of developmental toxicity in the absence of signs of marked maternal toxicity, or at around the same dose levels as other toxic effects but which are not a secondary non-specific consequence of the other toxic effects, but where are evidence is insufficient to place the substance in Category 2,
- Other relevant information.
-

The following symbols and specific risk phrases apply for category 3:
For substances which cause concern for human fertility: Xn; R62: Possible risk of impaired fertility
For substances which cause concern for humans owing to possible developmental toxic effects: Xn; R63: Possible risk of harm to the unborn child.

Table 13. Globally harmonised system of classification and labelling of chemicals

CATEGORY 1: Known or presumed human reproductive or developmental toxicant

This Category includes substances which are known to have produced an adverse effect on reproductive ability or capacity or on development in human or which there is evidence from animal studies, possibly supplemented with other information, to provide a strong presumption that the substance has the capacity to interfere with reproduction in humans. For regularly purposes, a substance can be further distinguished on the basis of whether the evidence for classification is primarily from human data (Category 1A or from animal data (Category 1B.

CATEGORY 1A: Known to have produced an adverse effect on reproductive ability or capacity or on development in humans

The placing of the substance in this category is largely based on evidence from humans.

CATEGORY 1B: Presumed to produce an adverse effect on reproductive ability or capacity or on development in humans

The placing of substance in this category is largely based on evidence from experimental animals. Data from animal studies should provide clear evidence of specific reproductive toxicity in the absence of other toxic effects, or if occurring together with other toxic effects the adverse effect on reproduction is considered not to be a secondary non-specific consequence of other toxic effects. However, when there is mechanistic information that raises doubt about the relevance of the effect for humans, classification in Category 2 may be more appropriate.

CATEGORY 2: Suspected human reproductive or developmental toxicant

This Category includes substance for which there is some evidence from human or experimental animals, - possibly supplemented with other information – of an adverse effect on reproductive ability or capacity, or on development, in the absence of other toxic effects, or if occurring together with other toxic effects the adverse effect on reproduction is considered not to be a secondary non-specific consequence of the other toxic effects, and where the evidence is not sufficiently convincing to place the substance in Category 1. For instance, deficiencies in the study may make the quality of evidence less convincing, and in view of this Category 2 could be the more appropriate classification.

5 References

Aafjes, J.H., Vels, J.M., Schenck, E. (1980). Fertility of rats with artificial oligozoospermia. J Reprod Fertil. 58, pp. 345-51.

Andersen, H. R., Andersson, A. M., Arnold, S. F., Autrup, H., Barfoed, M. et al. (1999). Comparison of short-term estrogenicity tests for identification of hormone-disrupting chemicals. Environ. Health Perspect. 107 Suppl 1, pp. 89-108.

Andersen, H. R., Vinggaard, A. M., Rasmussen, T. H., Gjermandsen, I. M., and Bonefeld-Jorgensen, E. C. (2002). Effects of currently used pesticides in assays for estrogenicity, androgenicity, and aromatase activity in vitro. Toxicol. Appl. Pharmacol. 179, pp. 1-12.

Arcadi, F. A., Costa, C., Imperatore, C., Marchese, A., Rapisarda, A., Salemi, M., Trimarchi, G. R., and Costa, G. (1998). Oral toxicity of bis (2-ethylhexyl) phthalate during pregnancy and suckling in the Long-Evans rat. Food Chem. Toxicol. 36, pp. 963-970.

Ashby, J. and Lefevre, P. A. (1997). The weanling male rat as an assay for endocrine disruption: preliminary observations. Regul. Toxicol. Pharmacol. 26, pp. 330-337.

Ashby, J. and Tinwell, H. (1998). Uterotrophic activity of bisphenol A in the immature rat. Environ. Health Perspect. 106, pp. 719-720.

Ashby, J. and Lefevre, P. A. (2000). The peripubertal male rat assay as an alternative to the Hershberger castrated male rat assay for the detection of anti-androgens, oestrogens and metabolic modulators. J. Appl. Toxicol. 20, pp. 35-47.

Bernal, J., Liewendahl, K., and Lamberg, B.A. (1985). Thyroid hormone receptors in fetal and hormone resistant tissues. Scand. J. Clin. Lab. Invest. 45, pp. 577-583.

Borch, J., Ladefoged, O., Hass, U., and Vinggaard, AM. (2003a) Steroidogenesis in fetal male rats is reduced by DEHP and DINP, but endocrine effects of DEHP are not modulated by DEHA in fetal, prepubertal and adult male rats. Reprod. Tox. Online: Doi:10.1016/j.reprotox. 2003.10.011.

Borch, J., Vinggaard, A.M., and Ladefoged, O. (2003b). The effect of combined exposure to di (2-ethylhexyl) phthlate and diisononyl phthlate on testosterone levels in foetal rat testis. Abstract. Reprod. Toxicol. 17, pp. 487-488.

Bowman, C. J., Barlow, N. J., Turner, K. J., Wallace, D. G., and Foster, P. M. (2003). Effects of in utero exposure to finasteride on androgen-dependent reproductive development in the male rat. Toxicol. Sci. 74, pp. 393-406.

Chapin, R. E., Delaney, J., Wang, Y., Lanning, L., Davis, B., Collins, B., Mintz, N., and Wolfe, G. (1999). The effects of 4-nonylphenol in rats: a multigeneration reproduction study. Toxicol. Sci. 52, pp. 80-91.

Clark, R. L., Antonello, J. M., Grossman, S. J., Wise, L. D., Anderson et al. (1990). External genitalia abnormalities in male rats exposed in utero to finasteride, a 5 alpha-reductase inhibitor. Teratology 42, pp. 91-100.

Clark, R. L., Anderson, C. A., Prahalada, S., Robertson, R. T., Lochry, E. A., Leonard, Y. M., Stevens, J. L., and Hoberman, A. M. (1993). Critical developmental periods for effects on male rat genitalia induced by finasteride, a 5 alpha-reductase inhibitor. Toxicol. Appl. Pharmacol. 119, pp. 34-40.

Cooke, P. S., Zhao, Y. D., and Hansen, L. G. (1996). Neonatal polychlorinated biphenyl treatment increases adult testis size and sperm production in the rat. Toxicol. Appl. Pharmacol. 136, pp. 112-117.

Course, J.F. Lindzey, J. Grandien, K. Gustafsson, J.Å., and Korach, K.S. (1997). Tissue distribution and quantitative analysis of estrogen receptor-α (ER- α) and estrogen receptor-β (ER- β) messenger RNA in the wild-type and ER α –knockout mouse. Endocrinology 138, pp. 4613-4648.

Creasy, D. M. (1997). Evaluation of testicular toxicity in safety evaluation studies: the appropriate use of spermatogenic staging. Toxicol. Pathol. 25, pp. 119-131.

Darnerud, P.O. (2003). Toxic effects of brominated flame retardants in man and in wildlife. Environ Int, 29, pp. 841-853.

EPA (United States Environmental Protections Agency), Endocrine Disrupter Screening and Testing Advisory Committee (EDSTAC) Final Report August 1998. http://www.epa.gov/scipoly/oscpendo/history/finalrpt.htm

EPA (2002). Draft detailed review paper on steroidogenesis screening assays and endocrine disruptors. May 2002.

Eriksson, P., Jakobsson, E., and Fredriksson, A. (2001). Brominated flame retardants: a novel class of developmental neurotoxicants in our environment? Environ. Health Perspec. 109, pp. 903-908.

EU 2003. Proposal for a Regulation of the European Parliament and of the Council concerning the Registration, Evaluation, Authorisation and Restrictions of Chemicals (REACH). Commission of the European Communities, draft proposal 29.10.03.

EU-Risk Assessment of 4-Nonylphenol (Branched) and Nonylphenol. Draft report September 1999.

EU-Risk Assessment of di(n-butyl) phthalate (DBP). Draft report June 2000.

EU-Risk Assessment of diisononyl phthalate (DINP). Draft report May 2001.

EU-Risk Assessment Report of 2,2-bis (4-hydroxyphenyl) propane (bisphenol A). Draft report July 2002.

EU-Risk Assessment of bis(2-ethylhexyl) phthalate (DEHP). Draft report August 2003.

Fail, P., Pearce, S. Anderson, S., Tyl, R., and Gray, L.E. (1995). Endocrine and reproductive toxicity of vinclozolin in male Long-Evans hooded rats. Toxicologist 15, pp. 293.

Fang, H., Tong, W., Perkins, R., Soto, A.M., Prechtl, N.V., and Sheehan, D.M. (2000). Quantitative comparison of in vitro assays for estrogenic activities. Environ. Health Perspec. 108, pp. 723-729.

Farwell, A.P. and Braverman, L.E. (1996). Inflammatory thyroid disorders. Otolaryngol. Clin. North Am. 29, pp. 541-556.

Goldman, J. M., Laws, S. C., Balchak, S. K., Cooper, R. L., and Kavlock, R. J. (2000). Endocrine-disrupting chemicals: prepubertal exposures and effects on sexual maturation and thyroid activity in the female rat. A focus on the EDSTAC recommendations. Crit. Rev. Toxicol. 30, pp. 135-196.

Gray, L. E., Jr., Ostby, J., Ferrell, J., Rehnberg, G., Linder, R., et al. (1989). A dose-response analysis of methoxychlor-induced alterations of reproductive development and function in the rat. Fundam. Appl. Toxicol. 12, pp. 92-108.

Gray, L. E. J., Kelce, W. R., Wiese, T., Tyl, R., Gaido, K., et al. (1997). Endocrine Screening Methods Workshop report: detection of estrogenic and androgenic hormonal and antihormonal activity for chemicals that act via receptor or steroidogenic enzyme mechanisms. Reprod. Toxicol. 11, pp. 719-750.

Gray, L. E. and Ostby, J. (1998). Effects of pesticides and toxic substances on behavioral and morphological reproductive development: endocrine versus nonendocrine mechanisms. Toxicol. Ind. Health 14, pp. 159-184.

Gray, L. E., Jr., Ostby, J., Monosson, E., and Kelce, W. R. (1999a). Environmental antiandrogens: low doses of the fungicide vinclozolin alter sexual differentiation of the male rat. Toxicol. Ind. Health 15, pp. 48-64.

Gray, L. E. J., Wolf, C., Lambright, C., Mann, P., Price, M., et al. (1999b). Administration of potentially antiandrogenic pesticides (procymidone, linuron, iprodione, chlozolinate, p,p'-DDE, and ketoconazole) and toxic substances (dibutyl- and diethylhexyl phthalate, PCB 169, and ethane dimethane sulphonate) during sexual differentiation produces diverse profiles of reproductive malformations in the male rat. Toxicol. Ind. Health 15, pp. 94-118.

Gray, L.E. Jr., Ostby, J., Wilson, V., Lambright, C., Bobseine, K., et al. (2002). Xenoendocrine disrupters-tiered screening and testing Filling key data gaps. Toxicology 181-182, pp. 371-382.

Haavisto, T. E., Adamsson, N. A., Myllymaki, S. A., Toppari, J., and Paranko, J. (2003). Effects of 4-tert-octylphenol, 4-tert-butylphenol, and diethylstilbestrol on prenatal testosterone surge in the rat. Reprod Toxicol 17, pp. 593-605.

Hallgren, S., Sinjari, T., Hakansson, H., and Darnerud, P.O. (2001). Effects of polybrominated diphenyl ethers (PBDEs) and polychlorinated biphenyls (PCBs) on thyroid hormone and vitamine A levels in rats and mice. Arch. Toxicol. 75, pp. 200-208.

Hallgren, S. and Darnerud, P.O. (2002). Polybrominated diphenyl ethers (PBDEs), polychlorinated biphenyls (PCBs) and chlorinated paraffins (CPs) in rats-testing interactions and mechanisms for thyroid hormone effects. Toxicology 177, pp. 227-243.

Harris, C. A., Henttu, P., Parker, M. G., and Sumpter, J. P. (1997). The estrogenic activity of phthalate esters in vitro. Environ. Health Perspect. 105, pp. 802-811.

Harvey, P.W. and Johnson, I. (2002). Approaches to assessment of toxicity data with endpoints related to endocrine disruption. J. Appl. Toxicol. 22, pp. 241-247.

Harvey, P.W. and Everett, D.J. (2003). The adrenal cortex and steroidogenesis as cellular and molecular targets for toxicity: Critical omissions from regulatory endocrine disrupter screening strategies for human health? J. Appl. Toxicol. 23, pp. 81-87.

Hass, U., Filinska, M., and Kledal, T.S.A. (2003). Antiandrogenic effects of Diisononyl phthalate in rats. Abstract. Reprod. Toxicol. 17, pp. 493-494.

Hellwig, J., van Ravenzwaay, B., Mayer, M., Gembardt, C. (2000). Pre- an dpostnatal oral toxicity of vinclozolin in Wistar and Long-Evans rats. Reg. Toxicol. Pharmacol. 32, pp. 42-50.

Hirsch, K. S., Weaver, D. E., Black, L. J., Falcone, J. F., and MacLusky, N. J. (1987). Inhibition of central nervous system aromatase activity: a mechanism for fenarimol-induced infertility in the male rat. Toxicol. Appl. Pharmacol. 91, pp. 235-245.

Hormonally active agents in the environment / Committee on Hormonally Active Agents in the Environment. Board on Environmental Studies and Toxicology, Commission on Life Sciences, National Research Council. (1999). Ed. By Knobil, E., National Academy Press, Washington D.C. ISBN 0-309-06419-8.

Hossaini, A., Dalgaard, M., Vinggaard, A. M., Pakarinen, P., and Larsen, J. J. (2003). Male reproductive effects of octylphenol and estradiol in Fischer and Wistar rats. Reprod. Toxicol.17, pp. 607-615.

Ishihara, A., Nishiyama, N., Sugiyama, S.and Yamauchi, K. (2003). The effect of endocrine disrupting chemicals on thyroid hormone binding to Japanese quail transthyretin and thyroid hormone receptor. Endocrinology 134, pp. 36-43.

Kanno, J., Onyon, L., Peddada, S., Ashby, J., Jacob, E., Owens, W. (2003). The OECD program to validate the rat uterotrophic bioassay. Phase 2: dose-response studies. Environ Health Perspect. Sep;111(12):1530-49.

Kelce, W. R., Monosson, E., Gamcsik, M. P., Laws, S. C., and Gray, L. E., Jr. (1994). Environmental hormone disrupters: evidence that vinclozolin developmental toxicity is mediated by antiandrogenic metabolites. Toxicol. Appl. Pharmacol. 126, pp. 276-285.

Kim, H.S., Han, S.Y., Kim, T.S., Kwack, S.J., Lee, R.D., Kim, I.Y., Seok, J.H., Lee, B.M., Yoo, S.D., Park, K.L. (2002). No androgenic/anti-androgenic effects of bisphenol-A in Hershberger assay using immature castrated rats. Toxicol Lett. Sep 5;135(1-2):111-23.

Korenbrot, C.C., Huhtaniemi, I.T., Weiner, R.I. (1977). Preputial separation as an external sign of pubertal development in the male rat. Biol. Reprod. 17, pp. 298-303.

Kuiper, G.G.J.M., Carlson, B., Grandj?ien, K.A.J., Haggblad, J., Nielson, S., and Gustafsson, J.Å. (1997). Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptor α and β. Endocrinology, 138, pp. 863-870.

Lambright, C., Ostby, J., Bobseine, K., Wilson, V., Hotchkiss, A.K., Mann, P.C., Gray, L.E. Jr. (2000). Cellular and molecular mechanisms of action of linuron: an antiandrogenic herbicide that produces reproductive malformations in male rats. Toxicol Sci. Aug;56(2):389-99.

Laws, S., Carey, S. A., Ferrell, J. M., Bodman, G. J., and Cooper, R. (2000). Estrogenic activity of octylphenol, nonylphenol, bisphenol A and methoxychlor in rats. Toxicol. Sci. 54, pp. 154-167.

Legler, J., Leonards, P., Spenkelink, A., and Murk, A.J. (2003). In vitro biomonitoring in polar extracts of solid phase matrices reveals the presence of unknown compounds with estrogenic activity. Ecotoxicology 12, pp. 239-249.

Markey, C.M, Michaelson, C.L, Veson, E.C, Sonnenschein, C., and Soto A.M. (2001). The mouse Uterotrophic bioassay: A reevaluation of its validity in assessing the estrogenicity of bisphenol A. Environ. Health. Perspec. 109, pp. 55-60.

Marty, M.S., Crissman, J.W., and Carney, E.W. (1999). Evaluation of the EDSTAC female pubertal assay in CD rats using 17beta-estradiol, steroid biosynthesis inhibitors, and a thyroid inhibitor. Toxicol. Sci. 52, pp. 269-277.

Marty, M. S., Crissman, J. W., and Carney, E. W. (2001). Evaluation of the male pubertal onset assay to detect testosterone and steroid biosynthesis inhibitors in CD rats. Toxicol Sci 60, pp. 285-295.

McIntyre, B. S., Barlow, N. J., and Foster, P. M. (2001). Androgen-mediated development in male rat offspring exposed to flutamide in utero: permanence and correlation of early postnatal changes in anogenital distance and nipple retention with malformations in androgen-dependent tissues. Toxicol. Sci. 62, pp. 236-249.

McIntyre, B. S., Barlow, N. J., and Foster, P. M. (2002). Male rats exposed to linuron in utero exhibit permanent changes in anogenital distance, nipple retention, and epididymal malformations that result in subsequent testicular atrophy. Toxicol. Sci. 65, pp. 62-70.

Monosson, E., Kelce, W. R., Lambright, C., Ostby, J., and Gray, L. E. J. (1999). Peripubertal exposure to the antiandrogenic fungicide, vinclozolin, delays puberty, inhibits the development of androgen-dependent tissues, and alters androgen receptor function in the male rat. Toxicol. Ind. Health 15, pp. 65-79.

MST (2001). Self-classification of dangerous substances, Danish Environmental Protection Agency, http://www.mst.dk/chemi/01050000.htm

Mylchreest, E., Sar, M., Cattley, R. C., and Foster, P. M. (1999). Disruption of androgen-regulated male reproductive development by di(n-butyl) phthalate during late gestation in rats is different from flutamide. Toxicol. Appl. Pharmacol. 156, pp. 81-95.

Mylchreest, E., Wallace, D. G., Cattley, R. C., and Foster, P. M. (2000). Dose-dependent alterations in androgen-regulated male reproductive development in rats exposed to Di(n-butyl) phthalate during late gestation. Toxicol. Sci. 55, pp. 143-151.

Mylchreest, E., Sar, M., Wallace, D. G., and Foster, P. M. (2002). Fetal testosterone insufficiency and abnormal proliferation of Leydig cells and gonocytes in rats exposed to di(n-butyl) phthalate. Reprod. Toxicol 16, pp. 19-28.

Nellemann, C., Vinggaard, A., Dalgaard, M., Hossaini, A., and Larsen, J. (2001). Quantification of antiandrogen effect determined by lightcycler technology. Toxicology 163, pp. 29-38.

Nellemann, C., Dalgaard, M., Lam, H. R., and Vinggaard, A. M. (2003). The combined effects of vinclozolin and procymidone do not deviate from expected additivity in vitro and in vivo. Toxicol. Sci. 71, pp. 251-262.

Newbold, R.R., Jefferson, W.N., Padilla-Banks, E., Walker, V.R., and Pena, D.S. (2001). Cell response endpoints enhance sesitivity of the immature mouse uterotrophic assay. Reprod. Toxicol. 15, pp. 245-252.

Nordic Council of Ministers. (2002). The Influence of chemicals in the food and the environment on the thyroid gland function. TemaNord: 520.

O'Connor, J. C., Frame, S. R., Davis, L. G., and Cook, J. C. (1999). Detection of thyroid toxicants in a tier I screening battery and alterations in thyroid endpoints over 28 days of exposure. Toxicol. Sci. 51, pp. 54-70.

O'Connor, J. C., Cook, J. C., Marty, M. S., Davis, L. G., Kaplan, A. M., and Carney, E. W. (2002a). Evaluation of Tier I screening approaches for detecting endocrine-active compounds (EACs). Crit. Rev. Toxicol. 32, pp. 521-549.

O'Connor, J. C., Frame, S. R., and Ladics, G. S. (2002b). Evaluation of a 15-day screening assay using intact male rats for identifying antiandrogens. Toxicol. Sci. 69, pp. 92-108.

O'Connor, J. C., Frame, S. R., and Ladics, G. S. (2002c). Evaluation of a 15-day screening assay using intact male rats for identifying steroid biosynthesis inhibitors and thyroid modulators. Toxicol. Sci. 69, pp. 79-91.

OECD (2002a) Detailed Review Paper, Appraisal of test methods for sex hormone disrupting chemicals. OECD Series on Testing and Assessment No. 21, ENV/JM/MONO(2002)8.

OECD (2002b), Draft summery report of the 6th meeting on the task force on endocrine disrupters testing and assessment (EDTA6). ENV/JM/TG/EDTA/M(2002)2/REV1.

OECD (2002c). OECD protocol for investigating the efficacy of the enhanced TG 407 test guideline (phase 2).

OECD (2003a), OECD Draft report of the validation of the rat Uterotrophic bioassay: Phase 2. testing of potent and weak oestrogen agonist by multiple laboratories. ENV/JM/TG/EDTA (2003)1.

OECD (2003b). OECD Draft report of the OECD validation of the rodent Hershberger bioassay: Phase 2. testing of androgen agonist, androgen antagonist and 5 alpha-reductase inhibitor in dose response studies by multiple laboratories. ENV/JM/TG/EDTA (2003)5.

OECD (2003c). OECD Draft of the validation of the phase two work results – Enhanced 407 studies – 20 May.

Ostby, J., Kelce, W. R., Lambright, C., Wolf, C. J., Mann, P., and Gray, L. E., Jr. (1999). The fungicide procymidone alters sexual differentiation in the male rat by acting as an androgen-receptor antagonist in vivo and in vitro. Toxicol. Ind. Health 15, pp. 80-93.

Owens, J.W and Ashby, J. (2002). Critical review and evaluation of the Uterotrophic bioassay for the identification of possible estrogen agonists and antagonists: In support of the validation of the OECD Uterotrophic protocols for the laboratory rodent. Crit. Rev. Toxicol. 32, pp. 445-520.

Parks, L. G., Ostby, J. S., Lambright, C. R., Abbott, B. D., Klinefelter, G. R., Barlow, N. J., and Gray, L. E. J. (2000). The plasticizer diethylhexyl phthalate induces malformations by decreasing fetal testosterone synthesis during sexual differentiation in the male Rat. Toxicol. Sci. 58, pp. 339-349.

Poon, R., Lecavalier, P., Mueller, R., Valli, V. E., Procter, B. G., and Chu, I. (1997). Subchronic oral toxicity of di-n-octyl phthalate and di (2-Ethylhexyl) phthalate in the rat. Food Chem. Toxicol. 35, pp. 225-239.

Shugrue, P.J., Lane M.V, Scrimo, P.J., and Merchenthaler, I. (1998). Comparative distribution of estrogen receptor-α (ER-α) and β (ER-β) mRNA in the rat pituitary, gonad, and reproductive tract. Steroids 63, pp. 498-504.

Simorangkir, D. R., de Kretser, D. M., and Wreford, N. G. (1995). Increased numbers of Sertoli and germ cells in adult rat testes induced by synergistic action of transient neonatal hypothyroidism and neonatal hemicastration. J. Reprod. Fertil. 104, pp. 207-213.

Spearow, J. L., Doemeny, P., Sera, R., Leffler, R., and Barkley, M. (1999). Genetic variation in susceptibility to endocrine disruption by estrogen in mice. Science 285, 1259-1261.

Soto, A. M., Sonnenschein, C., Chung, K. L., Fernandez, M. F., Olea, N., and Serrano, F. O. (1995). The E-SCREEN assay as a tool to identify estrogens: an update on estrogenic environmental pollutants. Environ. Health Perspect. 103 Suppl 7, pp. 113-122.

Stoker, T. E., Parks, L. G., Gray, L. E., and Cooper, R. L. (2000). Endocrine-disrupting chemicals: prepubertal exposures and effects on sexual maturation and thyroid function in the male rat. A focus on the EDSTAC recommendations. Endocrine Disrupter Screening and Testing Advisory Committee. Crit. Rev. Toxicol. 30, pp. 197-252.

Szelei, J., Jimenez, J., Soto, A.M., Luizzi, M.F., and Sonnenschein, C. (1997). Androgen-induced inhibition of proliferation in human breast cancer MCF7 cells transfected with androgen receptor. Endocrinology, pp. 1406-1412.

Térouanne, B., Tahiri, B., Georget, V., Belon, C., Poujol, N., Avances, C., et al. (2000). A stable prostatic bioluminescent cell line to investigate androgen and antiandrogen effects. Mol. Cell Endocrinol., pp. 39-49.

Timiras, P.S. and Nzekwe, E.U. (1989). Thyroid hormones and nervous system development. Biol. Neonate 55, pp. 376-385.

Tinwell, H., Joiner, R., Pate I., Soames, A., Foster, J., and Ashby, J. (2000). Uterotrophic activity of bisphenol A in the immature mouse. Reg. Toxicol. Pharmacol. 32, pp. 118-126.

Tyl, R. W., Myers, C. B., Marr, M. C., Brine, D. R., Fail, P. A., et al. (1999). Two-generation reproduction study with para-tert-octylphenol in rats. Regul. Toxicol. Pharmacol. 30, pp. 81-95.

Tyl, R. W., Myers, C. B., Marr, M. C., Thomas, B. F., Keimowitz, A. R., et al. (2002). Three-generation reproductive toxicity study of dietary bisphenol A in CD Sprague-Dawley rats. Toxicol. Sci. 68, pp. 121-146.

Vinggaard, A. M., Breinholt, V., and Larsen, J. C. (1999a). Screening of selected pesticides for oestrogen receptor activation in vitro. Food Addit. Contam. 16, pp. 533-542.

Vinggaard, A. M., Joergensen, E.C., and Larsen, J.C. (1999b). Rapid and sensitive reporter gene assays for detection of antiandrogenic and estrogenic effects of environmental chemicals. Toxicol. Appl. Pharmacol. 155, pp.150-160.

Vinggaard, A. M., Hnida, C., Breinholt, V., and Larsen, J. C. (2000). Screening of selected pesticides for inhibition of CYP19 aromatase activity in vitro. Toxicol. Vitr. 4, pp. 227-234.

Vinggaard, A. M., Nellemann, C., Dalgaard, M., Jorgensen, E. B., and Andersen, H. R. (2002). Antiandrogenic effects in vitro and in vivo of the fungicide prochloraz. Toxicol. Sci. 69, pp. 344-353.

Vinggard, A. M., Nellemann, C., Jarfelt, K., Dalgaard, M., Andersen H. R., Bonefelt-Jørgensen, E., and Hass, U. (2003). Antiandrogenic effects in vitro and in vivo of the fungicide procloraz. Abstract. Reprod. Toxicol. 17, pp.506.

Wade, J., Schlinger, B. A., Hodges, L., and Arnold, A. P. (1994). Fadrozole: a potent and specific inhibitor of aromatase in the zebra finch brain. Gen. Comp. Endocrinol. 94, pp. 53-61.

Walsh, L.P. and Stocco, D.M. (2000). Effects of Lindane on steroidogenesis acute regulatory protein expression. Biol. Reprod. 63, pp. 1024-1033.

Wilson, V.S., Bobseine, K., Lambright, C.R., and Gray, L.E.Jr. (2002). A novel cell line, MDA-kb2, that stably expresses an androgen- and glucocorticoid-responsive reporter for the detection of hormone receptor agonists and antagonists. Toxicol. Sci. 66, pp. 69-81.

Yamasaki, K., Takeyoshi, M., Yakabe, Y., Sawaki M, et al. (2002a). Comparison of reporter gene assays and immature rat uterotrophic assay of twenty- three chemicals. Toxicology 170, pp. 21-30.

Yamasaki, K., Sawaki, M., Noda S., Imatanaka, N., and Takatsuki, M. (2002b). Subcutane oral toxicity study of ethinylestradiol and bisphenol A, based on the draft protocol for the 'Enhanced OECD Test Guideline no. 407. Arch. Toxicol. 76, pp. 65-74.

Zhou, T., Taylor, M.M., DeVito, M.J., and Crofton, K.M. (2002). Developmental exposure to brominated diphenyl ethers results in thyroid hormone disruption. Toxicol. Sci. 66, pp. 105-116.

Zoeller, R.T. (2003). Challenges confronting risk analysis of potential thyroid toxicants. Risk analysis 23, pp. 143-162.

6 Appendix

Summary of endpoints in assays at level four

Study/endpoints	**Pubertal male assay[a]**	**Pubertal female assay[b]**	**Intact male assay[c]**	**Enhanced TG 407[d]**
Preferred species	Rat	Rat	Rat	Rat
Growth	X	X		X
Age at balano preputial separation	X			
Serum T_4 and TSH	X	X	X	X
Thyroid histology	X	X	X	X
Seminal vesicle weight	X		X	X
Seminal vesicle histology				X
Ventral prostate weight	X		X	X
Ventral prostate histology				X
Levator ani/bulbocavernosus muscle weight	X			
Testis weight	X		X	X
Testis histology	X		X	X
Epididymal weight	X		X	X
Epididymal histology	X		X	X
Age at vaginal opening		X		X
Vaginal cytology		X		X
Vagina histology				X
Ovarian weight		X		X
Ovarian histology		X		X
Uterine weight		X		X
Uterine histology		X		X
Liver weight	O	X	X	X
Liver histology				X
Mammary glands weight and histology				X
Kidney weight	O	X		X
Kidney histology				X
Pituitary weight	O	X		X
Pituitary histology				X
Adrenal weight	O	X		X
Adrenal histology				X
Thymus, spleen, pancreas, brain and heart weight and histology				X
Haematology				X
Clinical biochemistry				X
Serum testosterone	O			
Estradiol	O		X	
LH	O		X	
Prolactin	O		X	
T3	O		X	X
Ex vivo testis and pituitary hormone production	O			
Hypothalamic neurotransmitter levels	O			
Sperm concentration and morphology in cauda epididymis			X?	X

A Immature (23-54 days of age) intact male rat protocol to evaluate pubertal development and thyroid function

B Immature (21-43 days of age) intact female rat protocol to evaluate pubertal development and thyroid function
C Intact male rats (70-90 days of age) are used.
D Intact male and females (49 days of age) are used.
X Guideline specifies parameter as endpoint that requires routine assessment
O Guideline specifies parameter as possible (optional) endpoint on a study-specific basis

Summary of relevant endpoints in reproductive toxicity studies (modified from OECD 2001a)

Study/endpoints	**415**	**416**	**421**	**422**	**426**	**414**
Preferred species	Rat/mouse	Rat/mouse	Rat	Rat	Rat	Rat/rabbit
Adult						
General behaviour	X	X	X	X	X	X
Gonadal weight/size	A	X	X-M, A-F	X-M, A-F	A	A
Gonadal gross pathological appearance	X	X	X	X	X	X
Gonadal histopathology	O	X	X	X	X?	X
Accessory reproductive organ weight/size	A	X	X	X	X?	A (X)
Accessory reproductive organ gross pathological appearance	X	X	X	X	X	X
Accessory reproductive organ histopathology	O	O (X)	X-M, A-F	X	X?	A
Accessory reproductive organ secretory product formation		A	A	A	A	A
Blood levels of sex hormones	A	A	A	X?	A-F	
Spermatogenesis (detailed histopathological assessment)	O	O (X)	A	A	A-F	
Sperm count/quality assessment	A	X	A	A		
Sperm motility/morphology	A	X	A	A		
Oestrus cyclicity	O	X	A	A	A	
Time to mating	X	X	X	X		
Mating/sexual behaviour	X?	X?	X?			
Gestation length	X	X	X	X	X	
Corpora lutea count				O		X
Litter pre/post implantation losses and abortions	X	X	X	X	X	X
Premature delivery	X	X	X	X	X	X
Dystocia	X	X	X	X	X	
Fecundity (size/no. of litters)	X	X	X	X	X	X
Reproductive life span						
Maternal/lactational ability	X	X	X	X	X	
Offspring						
General behaviour		O			X	
Sex ratio	X	X	X	X	X	X
Sex differentiation	X	X	X	X	X	X?
Foetal weight						X
Sexual maturation		X			X	
Growth rate	X	X	X	X	X	
Learning and memory/sexual dimorphic behaviour	A	O			X	
Gonadal weight/size	A	X			X?	
Gonadal gross pathological appearance	X	X	X	X	X?	
Accessory reproductive organ weight/size	A	X	A	A	A	

Accessory reproductive organ gross pathological appearance	X	X	X	X	X	
Accessory reproductive organ histopathology	A	X	A	A	X?	
Accessory reproductive organ secretory product formation		A			A	
Blood levels of sex hormones		A			A	
Spermatogenesis (detailed histopathological assessment)		X			A	
Sperm count/quality assessment		A			A	
Sperm motility/morphology		X?			A	
Oestrus cyclicity		X			A	
Time to mating		X				
Mating/sexual behaviour		X				
Gestation length		X				
F2 litter pre/post implantation losses and abortion		X				
Premature delivery		X				
Dystocia		X				
F2 growth, development etc		X				

X Guideline specifies parameter as endpoint that requires routine assessment
O Guideline specifies parameter as possible (optional) endpoint on a study-specific basis
A Endpoint which could be included in study design without significant changes of design

+