Methods in Pharmacology and Toxicology

Springer Protocols

Ali S. Faqi *Editor*

Developmental and Reproductive Toxicology

Methods in Pharmacology and Toxicology

Series Editor
Y. James Kang
University of Louisville
School of Medicine
Prospect, Kentucky, USA

For further volumes:
http://www.springer.com/series/7653

Developmental and Reproductive Toxicology

Edited by

Ali S. Faqi

MPI Research
Mattawan, MI, USA

Editor
Ali S. Faqi
MPI Research
Mattawan, MI, USA

ISSN 1557-2153 ISSN 1940-6053 (electronic)
Methods in Pharmacology and Toxicology
ISBN 978-1-4939-7206-7 ISBN 978-1-4939-7208-1 (eBook)
DOI 10.1007/978-1-4939-7208-1

Library of Congress Control Number: 2017946540

© Springer Science+Business Media LLC 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Dedication

This volume is dedicated to Somalia and to the suffering people of Somalia. In addition, I dedicate this to my parents, my elder brother Abdulqadir Said Faqi, my dear Yasmine Allas, and my colleague Lisa Heimsath.

Preface

Developmental toxicity is defined as the study of adverse effects on the developing organism that may result from exposure to drugs/chemicals prior to conception (either parent), during prenatal development, or postnatally to the time of sexual maturation.

The thalidomide disaster is widely believed to be the catalyst that promoted regulatory agencies around the world, including the US FDA, to initiate requirements for new drugs to be thoroughly tested in animals prior to being sold in the marketplace.

At that time, developmental toxicity studies conducted in animals were inappropriately designed and insufficient to detect a teratogenic signal.

We currently rely on animal testing to predict the potential for drugs or chemicals to cause developmental toxicity in humans. Rodents (rats and mice) and rabbits are the most relevant species used in developmental toxicity testing, dogs and minipigs are rarely used, and nonhuman primates may be used for biologics, especially for monoclonal antibodies.

Manifestation of developmental and reproductive toxicity may include adverse effects on onset of puberty, gamete production and transport, reproductive cycle normality, sexual behavior, fertility, gestation, parturition, lactation, structural abnormalities, premature reproductive senescence, and modifications of other functions that are dependent on the integrity of the reproductive systems.

Evaluation of developmental and reproductive toxicology endpoints is an integral part of the safety assessment process for compounds with potential use in women of childbearing age or females that might be exposed during pregnancy as well as men of reproductive potential.

This volume covers metabolism and drug-drug interactions during pregnancy, critical periods of developmental toxicology, in vivo and alternative methods to assess potential developmental toxicity for drugs and chemicals, and effects of chemicals on testes and mammary glands. The in vivo assessments are guideline-driven and are required for submissions for product approval.

On the other hand, alternative methods for developmental toxicity testing have been sought because of the pressure to reduce the number of animals used in health research. Alternative in vitro methods include cell cultures, zebra fish, c-elegans, organ cultures, and embryo cultures and embryonic stem cells. These test systems can provide invaluable information and decrease the number of animals used in studies. The design of in vitro alternatives with good predictivity of in vivo effects is challenging, as embryo-fetal development is a continuous process of a precisely orchestrated sequence of events and any alternative assay in the field of developmental toxicity represents only part of the complexity of the whole developing conceptus and its maternal environment. Currently, the alternative methods are not used for regulatory submissions but mainly for screening and mechanistic studies.

Finally, I would like to thank all the authors/coauthors for their hard work and timely contributions. Likewise, I would like to extend my sincerest thanks and appreciation to David Casey and the entire Springer publishing team who worked tirelessly in the publication of this volume.

Mattawan, MI, USA ***Ali S. Faqi***

Contents

Preface . . . . *vii*
Contributors . . . . *xi*

Metabolism and Drug–Drug Interaction in Pregnant Mother/Placenta/Fetus 1
Ali S. Faqi and Karsten A. Holm

Critical Periods of Development in Teratology 17
Padmanabhan Rengasamy

Nonclinical Safety Assessment of Developmental and Reproductive Toxicology: Considerations for Conducting Fertility, Embryo–Fetal Development, and Prenatal and Postnatal Developmental Toxicology Studies 43
Kevin H. Denny and Ali S. Faqi

A Developmental and Reproductive Toxicology Program for Chemical Registration 117
Kamin J. Johnson, Bethany R. Hannas, Sue Marty, Carol L. Zablotny, Jessica LaRocca, Nicholas Ball, and Amanda K. Andrus

The Influence of Environmental Contaminants and Lifestyle on Testicular Damage and Male Fertility 185
Sellappan Selvaraju, Meritxell Jodar, and Stephen A. Krawetz

Effects of Chemicals on Mammary Gland Development 205
Adam J. Filgo and Ali S. Faqi

Alternative Methods Used to Assess Potential Embryo-Fetal Developmental Risk of Pharmaceuticals 235
Maia L. Green, Joshua T. Gamse, and Kimberly C. Brannen

Using the Alternative Model *C. elegans* in Reproductive and Developmental Toxicology Studies 261
Daniel W. Ferreira, Yichang Chen, and Patrick Allard

High-Throughput Screens for Embryonic Stem Cells: Stress-Forced Potency-Stemness Loss Enables Toxicological Assays 279
Quanwen Li, Yu Yang, Erica Louden, Elizabeth E. Puscheck, and Daniel A. Rappolee

Index *313*

Contributors

PATRICK ALLARD • *Department of Environmental Health Sciences, University of California, Los Angeles, Los Angeles, CA, USA; Molecular Toxicology, Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, USA; Institute for Society & Genetics, University of California, Los Angeles, Los Angeles, CA, USA*

AMANDA K. ANDRUS • *Toxicology and Environmental Research & Consulting, The Dow Chemical Company, Midland, MI, USA*

NICHOLAS BALL • *Toxicology and Environmental Research & Consulting, Dow Europe GmbH, Horgen, Switzerland*

KIMBERLY C. BRANNEN • *Safety Assessment and Laboratory Animal Resources, Merck & Co., Inc., West Point, PA, USA*

YICHANG CHEN • *Department of Environmental Health Sciences, University of California, Los Angeles, Los Angeles, CA, USA; Molecular Toxicology, Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, USA*

KEVIN H. DENNY • *Teva Branded Pharmaceutical Products Research and Development, Inc., West Chester, PA, USA*

ALI S. FAQI • *MPI Research, Mattawan, MI, USA*

DANIEL W. FERREIRA • *Department of Environmental Health Sciences, University of California, Los Angeles, Los Angeles, CA, USA*

ADAM J. FILGO • *MPI Research, Mattawan, MI, USA*

JOSHUA T. GAMSE • *Division of Drug Safety Assessment, Department of Reproductive Toxicology, Bristol-Myers Squibb, New Brunswick, NJ, USA*

MAIA L. GREEN • *Safety Assessment and Laboratory Animal Resources, Merck & Co., Inc., West Point, PA, USA; MRL, West Point, PA, USA*

BETHANY R. HANNAS • *Toxicology and Environmental Research & Consulting, The Dow Chemical Company, Midland, MI, USA*

KARSTEN A. HOLM • *Alliance Pharma, Inc., Malvern, PA, USA*

MERITXELL JODAR • *Department of Obstetrics and Gynecology, Center for Molecular Medicine and Genetics, C.S. Mott Center for Human Growth and Development, Wayne State University School of Medicine, Detroit, MI, USA*

KAMIN J. JOHNSON • *Toxicology and Environmental Research & Consulting, The Dow Chemical Company, Midland, MI, USA*

STEPHEN A. KRAWETZ • *Department of Obstetrics and Gynecology, Center for Molecular Medicine and Genetics, C.S. Mott Center for Human Growth and Development, Wayne State University School of Medicine, Detroit, MI, USA*

JESSICA LAROCCA • *Human Health Assessment, Dow AgroSciences LLC, Indianapolis, IN, USA*

QUANWEN LI • *Department of Obstetrics and Gynecology, C.S. Mott Center for Human Growth and Development, Wayne State University School of Medicine, Detroit, MI, USA*

ERICA LOUDEN • *Department of Obstetrics and Gynecology, C.S. Mott Center for Human Growth and Development, Wayne State University School of Medicine, Detroit, MI, USA*

SUE MARTY • *Toxicology and Environmental Research & Consulting, The Dow Chemical Company, Midland, MI, USA*

ELIZABETH E. PUSCHECK • *Department of Obstetrics/Gynecology, Reproductive Sciences Program, University School of Medicine, Detroit, MI, USA*

DANIEL A. RAPPOLEE • *Department of Obstetrics and Gynecology, C.S. Mott Center for Human Growth and Development, Wayne State University School of Medicine, Detroit, MI, USA; Department of Physiology, Reproductive Sciences Program, Wayne State University School of Medicine, Detroit, MI, USA; Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI, USA; Institutes for Environmental Health Science, Wayne State University School of Medicine, Detroit, MI, USA; Department of Biology, University of Windsor, Windsor, ON, Canada*

PADMANAHBAN RENGASAMY • *Department of Medical Education, California Northstate University College of Medicine, Elk Grove, CA, USA*

SELLAPPAN SELVARAJU • *Department of Obstetrics and Gynecology, Center for Molecular Medicine and Genetics, C.S. Mott Center for Human Growth and Development, Wayne State University School of Medicine, Detroit, MI, USA; National Institute of Animal Nutrition and Physiology, Bangalore, India*

YU YANG • *Department of Obstetrics and Gynecology, C.S. Mott Center for Human Growth and Development, Wayne State University School of Medicine, Detroit, MI, USA; Department of Physiology, C.S. Mott Center for Human Growth and Development, Wayne State University School of Medicine, Detroit, MI, USA*

CAROL L. ZABLOTNY • *Toxicology and Environmental Research & Consulting, The Dow Chemical Company, Midland, MI, USA*

Methods in Pharmacology and Toxicology (2017) 1–16
DOI 10.1007/7653_2016_64
© Springer Science+Business Media New York 2016
Published online: 29 May 2016

Metabolism and Drug–Drug Interaction in Pregnant Mother/Placenta/Fetus

Ali S. Faqi and Karsten A. Holm

Abstract

The pregnant woman and the presence of the fetus pose many challenges for proper and effective drug administration. The variety of physiological changes that takes place during pregnancy coupled together with the variety in the responses of the cytochrome P450 enzymes in terms of induction and inhibition as well as the presence of polymorphic forms which may be present and the influence of the drug transporters make predicting the pharmacokinetics and pharmacodynamics of any given drug difficult. Treatment and dosage during pregnancy and lactation with drugs such as antibiotics, antivirals, antiepileptic, anticancer, and antipsychotic medications all need to be evaluated carefully to minimize the occurrence of adverse effects due to possible excessive exposure or a lack of efficacy due to possible underexposure. In addition, as more literature data becomes available about the role of efflux transporters such as Pgp, BCRP, and MRP3 and uptake transporters OCT3 and OCTN1 in pregnancy and in the fetus with prescribed medications this information will need to be used in the evaluation. Therefore, for drugs with a narrow therapeutic window or those with marked pharmacologic or toxicological outcomes that are also cleared predominantly by a single CYP450 or handled by a single transporter, the need for systemic monitoring of plasma concentration to monitor exposure is warranted, at least during the initial days of starting a medication.

Keywords: Drug–drug interaction, Drug disposition, Pregnancy, Lactation

1 Introduction

Sixty-five percent of pregnant women in the USA take one or more drugs during their pregnancy. This does not include dietary supplements or vitamins [1].

During pregnancy the effect of drugs may very differently due to several pregnancy-induced changes in drug disposition thus making the efficacy and toxicity of drugs used by pregnant women difficult to predict. Two factors influencing both drug efficacy and disposition are the drug-metabolizing enzymes and drug transporters. Among the more influential drug-metabolizing enzymes are the enzymes of the cytochrome P450 family (CYP450s) as these enzymes are centrally involved in the disposition of the majority of drugs, they exist in many genetic variations and are regulated by multiple mechanisms allowing for their induction and/or inhibition.

In vivo studies have shown that the activity of several hepatic cytochrome P450 enzymes, such as CYP2D6 and CYP3A4, is increased during pregnancy, whereas the activity of others, such as CYP1A2, is decreased. Likewise, the activity of some renal transporters, including organic cation transporter and P-glycoprotein, appears to be increased during pregnancy [2].

The multiple forms of the CYP450s and their activities have been described in detail in numerous reports and reviews [3–5] and are touched on briefly in this chapter. The drug transporters are a newer area of intense research into the complexity of factors influencing drug pharmacology and disposition due to their integral role in drug absorption, exposure, elimination and thus an additional source of drug–drug interactions. The pharmacokinetic changes due to genetic polymorphisms and drug–drug interactions involving transporters can often have a direct impact on the therapeutic safety and efficacy of many important drugs [6]. The transporters studied and described to date are primarily from the major organs involved in drug uptake and disposition such as the GI tract, liver, kidney, and brain as described by Borst et al. [7].

The P450 metabolic pathways through their actions on drugs, endogenous compounds and concomitantly administered medications are a major source of drug–drug, drug–diet, and drug–disease/condition interactions; consequently, functional variability in this complex system can have pronounced consequences in suboptimal therapeutic response or enhanced toxicity [8]. The regulation of the numerous CYP450s is becoming better understood as research in this area continues. The genetic factors and physiological processes controlling CYP450 levels and their induction/inhibition properties are well documented [9]. Additionally, the effects of various nutritional and a disease state such as fasting, diabetes, malnutrition, and alcohol abuse on these systems has been examined and the changes in CYP450s have been discussed [8].

However, there is not much information available about changes in the CYP450s and transporter systems during pregnancy and lactation in the human female as this is an area of more recent investigation and just beginning to be understood. Understanding the physiological and biochemical factors that change in the human female during pregnancy and how they influence pharmacokinetic factors, the CYP450s and transporters is important as medication during pregnancy is common, but specific information about the changes in how these medicines are processed as a result of pregnancy or what the drug exposure is to the mother, placenta, and fetus is not fully known or understood. An accurate understanding of the pharmacokinetics and metabolism of drugs during pregnancy is essential for the safe and optimal drug therapy for the mother and fetus, thus, it is important to have a full understanding of how pregnancy influences drug disposition

factors for better therapeutic outcomes and better predictions of the pharmacokinetic changes of drugs and their effects in pregnant women as well as the fetus. This will allow better prediction of pharmacokinetic changes of drugs in pregnant women. Therefore, the goal of this review is to present what is known about these enzyme and transporter systems and how they change in women during pregnancy and lactation, in the placenta and in the fetus. In addition, the review also discusses any known drug–drug interactions in the pregnant mother/placenta and the fetus.

1.1 Drug Disposition Changes During Pregnancy and Lactation

The pharmacokinetics of various drugs may be profoundly altered during different stages of pregnancy, parturition, and lactation due to numerous physiological and biochemical changes that takes place during pregnancy. During pregnancy the physiological changes include plasma volume expansion and increases in extracellular fluid space and total body water; decreased plasma albumin concentration; a compensated respiratory alkalosis; increased cardiac output with regional blood flow changes; increased renal blood flow associated with increased glomerular filtration; changes in hepatic drug-metabolizing enzymes; and reduction in intestinal motility, increased glomerular filtration rate, and reduced plasma albumin concentration [10]. The increases in plasma volume and total body water may increase the volume of distribution and thereby increase the dose requirements that are necessary to sustain therapeutic drug levels [11].

These changes begin in early gestation but are most pronounced in the third trimester of pregnancy. More maternal physiologic changes occur intrapartum with some normalizing themselves within 24 h of delivery, while others are more prolonged only returning to normal some 12 weeks postpartum [12]. All these changes modify drug distribution, metabolism, and elimination. As a result, the exposure and disposition of many medicines may be altered during pregnancy and the resulting clinical efficacy and toxicity of these drugs can be difficult to predict or can lead to serious side effects. An increase in body weight during pregnancy may result in a decrease in dose per kilogram and thus a potential for a significant lowering of a drug's steady state concentration and thus possible suboptimal treatment.

Additionally, gastrointestinal absorption or bioavailability of drugs may vary due to changes in gastric secretion and motility.

Multiple hemodynamic changes such as an increase in cardiac output, blood volume, and renal plasma flow may affect drug disposition and elimination [13]; these changes in pharmacokinetic parameters should be considered when dosing antiarrhythmic agents in pregnant women [14]. Absorption of drugs may be decreased by nausea and vomiting associated with pregnancy, especially in the first trimester [15]. There are also increases in hormone levels, particularly estrogen, progesterone, placental growth

hormone, and prolactin which have multiple effects particularly on the drug-metabolizing enzymes. One possible effect of the hormonal change is on absorption, the increased plasma progesterone concentrations during pregnancy corresponds to decreases in gastrointestinal motility, with associated prolonged gastric emptying and intestinal transit times which may lead to delayed drug absorption and reduced peak concentrations [16]. Indeed an in-depth understanding in hormonal regulatory mechanisms is warranted for systematic understanding and prediction of the changes in hepatic drug metabolism during pregnancy [17].

Additional absorption changes for weak acid and basic drugs are due to increased gastric pH due to reduced gastric acid secretion which may affect the ionization and absorption of week acids and bases [18]. The increase in blood and total body water volumes can alter the volume of distribution for various drugs. These changes may affect drug disposition and elimination, and can cause an increase or decrease in the terminal elimination half-life of drugs.

1.2 Enzyme Influences During Pregnancy and Lactation

The enzymes of the cytochrome P450 family (CYP450s) have a central role in the pharmacokinetics and metabolism of most medicines in clinical use today. They have been extensively studied ever since there discovery in the 1950s. The majority of CYP enzymes are found in the liver, although other organs such as the gastrointestinal mucosa, skin, lung, brain, and kidney also have significant CYP expression and functional activity [19].

Pollutants and toxicants passing from the mother to the fetus may damage developing organ systems. The human fetal liver is both a potential target organ and a critical defense against exposure to such chemicals. Exposure of the fetus to pollutants/toxicants is associated with significantly altered transcript expression, with the more marked response in the male potentially affecting levels of endogenous factors involved in fetal growth [20].

The activities and nomenclature have been better defined over the last 20 years. Although there are more than 100 CYP genes in humans, there are only about 10 gene products that are important to monitor in preclinical and clinical development for potential drug–drug interactions as reported by Huang et al. [21], namely, CYPs 1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, 3A4, and 3A5. These CYPs have the potential for not only inhibition and induction but also genetic polymorphisms that can produce clinically important outcomes. Several of the drug-metabolizing enzymes are polymorphic, having more than one variant of the gene.

A prospective cohort study of 293 women, who delivered singleton live births in Sapporo, Japan, was conducted to estimate the effects of maternal smoking and genetic polymorphisms on infant birth weight and length. Birth weight and length were

significantly lower among infants born to continuously smoking women having the aryl hydrocarbon receptor (*AhR)* wild type genotype the *CYP1A1* variant genotype or the *GSTM1* null genotype indicating that maternal smoking in combination with maternal *AhR*, *CYP1A1*, and *GSTM1* genetic polymorphisms may adversely affect infant birth size [22].

The CYP2D6 enzyme is perhaps the most widely recognized polymorphic enzyme with a recessive poor metabolizer (PM) phenotype resulting when individuals carry two null alleles, yielding either a completely metabolically inactive protein, or no protein [23]. Although the CYP isozymes generally have similar structural and overlapping functional properties, each form has key structural differences resulting in distinct functional properties creating a distinct pattern of metabolic reactions for given substrates. Although CYP2D6 mRNA is detectable in the fetus, however, CYP2D6 protein expression remains mostly undetectable during pregnancy. The CYP2D6 protein concentration rises only a few days after birth [24], but remains low during the first month of life (about 20 % of adult's levels [25]. During the lactation (newborn) period a low level of CYP2D6 activity occurs, independent of genotype that functionally results in all new born being poor metabolizers, as a result, clearance of CYP2D6 substrates are expected to be low for almost all infants. In order to prevent drug accumulation or toxicity individualization of dosing is necessary in infants [23]. The variability of CYP2D6 activities in infants older than 1–2 weeks was largely found to be related to genetic variability [26].

With the additional dimension of genetic polymorphisms there is an increased basis for interindividual differences in the pharmacologic efficacy and side effects of drugs as well as their toxicological and carcinogenic potential. The variability associated with the CYP450 enzymes between individuals result in marked differences in responses when the same drug and the dose are administered to different individuals [12].

In addition to the genetic polymorphism found in CYP2D6, this has been also found in the CYP2C family, specifically 2C8, 2C9, and 2C19 [27].

The expression of CYP3A4 is low during pregnancy and at birth. However, it is the primary hepatic CYP expressed postnatally and is involved in the metabolism of over 75 % of commonly used drugs [23]. CYP3A7 is the major CYP isoform detected in embryonic, fetal, and newborn liver with a shift between the CYP3A7 and CYP3A4 occurring after birth [25].

Loss of consciousness was reported in neonates receiving coadministration of erythromycin (an inhibitor of CYP 3A4) with midazolam [28]. Drug–drug interactions can occur if a drug acts as an inducer or inhibitor of a CYP450 enzyme and significantly alters the function of that enzyme or if an individual has a polymorphic variant form of a CYP450 enzyme. The degree of induction or

inhibition of CYP 3A4 might be influenced by the developmental changes which could further enhance the drug–drug interactions in an immature system.

Regarding changes found in the CYP450 system in pregnant adult women, it has been shown that the changes are variable and affect only a few of the CYP450 enzymes. It has been demonstrated that the activity of the CYP2C subfamily, CYP2D6 and CYP3A4 enzymes increase, while in contrast, the activity of CYP1A2 decreases [29, 30]. The study by Wadelius et al. [30] on CYP2D6 activity involved 17 pregnant women phenotyped into 3 groups with 4 as poor metabolizers, seven as heterozygous extensive metabolizers and six as homozygous extensive metabolizers with dextromethorphan in late pregnancy and 7–11 weeks after parturition. During pregnancy, the metabolic ratio of dextromethorphan and dextrorphan was significantly reduced ($p = 0.0015$) in the homozygous and heterozygous extensive metabolizers, consistent with increased CYP2D6 activity. In contrast, the poor metabolizers showed an increased metabolic ratio during pregnancy. This study finding was consistent with a previous study finding which found a marked increase in the metabolism of the CYP2D6 substrate metoprolol during pregnancy. Because both studies found an increase in CYP2D6 activity during pregnancy, it was suspected that pregnancy somehow causes the induction of the CYP2D6 enzyme. The findings of Davis et al. and Wadelius et al. [29, 30] have been recently confirmed in a study by Tracy et al. [31] which also showed increases in CYP2D6 and CYP3A4 activity and a decrease in CYP1A2 activity. In this study 25 subjects completed the study conducted at several stages of pregnancy, 14–18 weeks, 24–28 weeks and 36–40 weeks, and again at 6–8 weeks after delivery. The enzyme activity results from the 3 phases of pregnancy were compared with the postpartum period. It was found that CYP1A2 activity decreased progressively during the pregnancy relative to the postpartum period with activity reductions of 33, 48 and 65 at 14–18 weeks, 24–28 weeks, and 36–40 weeks, respectively. CYP2D6 activity increased over the course of the pregnancy relative to the postpartum period with increases of 26, 35 and 48 % at 14–18 weeks, 24–28 weeks, and 36–40 weeks, respectively. CYP3A4 activity increased consistently and similarly at each phase relative to the postpartum period with increased activity between 35 and 38 %. Thus, pregnancy can cause opposing actions on the CYP450 system with increases in CYP2D6 and CYP3A4 activity and a decrease in CYP1A2 activity [31]. Recently, increased Cyp3A4 expression; unchanged Cyp2A5 expression and decreased Cyp1A2, Cyp2C37, Cyp2D22, Cyp2E1, and Cyp3A11 was reported in mice during pregnancy Also expression of CYPD22 and CYP2 E1 isoforms correlated with that of peroxisome proliferator-activated receptor PPARα in the mouse livers, suggesting potential involvement of PPARα in downregulation of the

P450 expression during pregnancy [32]. In addition, they found that the expression of Cyp2D22 and Cyp2E1 isoforms directly correlated with that of peroxisome proliferator-activated receptor (PPAR) α in the mouse livers, which led them to suggest potential involvement of PPARα in downregulation of the P450 expression during pregnancy. It is fair therefore to conclude that any dosing adjustment during pregnancy will depend on the medication and the enzyme involved in its metabolism.

Another important aspect is the formation of toxic metabolites that could lead to birth effects. For example, a genetic defect in arene oxide detoxification seems to increase the risk of the baby having major birth defects in epileptic women treated with phenytoin [33]. Shanks et al. [34] developed a murine embryo culture model to study the potential contribution of enzymatic bioactivation to the teratogenicity of phenytoin. Their result suggest that the embryo can enzymatically bioactivate embryotoxically significant amounts of phenytoin, and that bioactivation and embryotoxicity is further enhanced, by an exogenous P-450 system, implicating a possible maternal contribution to phenytoin teratogenicity. A literature review performed on pharmacogenetics of drug induced birth defects found that direct relationship between pharmacogenetics and drug-induced birth defects exits for folate metabolism, oxidative stress caused by phenytoin exposure and drug transporters in the placenta [35].

It has been also been suggested that an increased metabolic conversion of valproate (VPA) to its toxic metabolites including 2-propyl-4-pentenoic acid (4-en) is involved in the mechanism of VPA teratogenicity at higher doses and concentrations [36].

The impact of development and CYP2C9 polymorphisms on neonatal therapeutics can be explained by the interindividual variability for AUC values reported when ibuprofen and indomethacin are used for treatment of ductus arteriosus in neonates. Although indomethacin had a higher volume of distribution in the very preterm baby; clearance from the blood stream occurs more quickly in babies more than 1–2 weeks. In addition markedly longer half-life was observed for Ibuprofen [37].

Although the impact of ontogeny for Phase II enzymes is less studied than phase I enzymes; however, the understanding of their developmental profiles is essential to recognizing the acquisition of metabolic competence in the neonate and its potential therapeutic implications [38]. Glutathione S-transferase (GST) A1 and A2 were identified in human fetal liver tissues during gestation as early as 10 weeks gestational age with adult levels not reached until 1–2 years. For GSTP1, the fetal kidney expression pattern at less than 35 weeks gestational age was similar to that observed for GSTA1/A2. In fetal tissue greater than 35 weeks of age, expression was restricted to collecting tubules and the distal loop of Henle [38]. The presence of GST isoforms in urinary epithelia, digestive tract,

and respiratory tract highlights the importance of GST in detoxification reactions at a very early age and suggests that the embryo is capable of metabolizing drugs [23]. Maternal exposure to these chemicals that induce GST including non-nutrient xenobiotics found in vegetables and citrus fruits have the potential to alter drug metabolism during pregnancy and lactation [23]. The tragedy of Gray Baby Syndrome was the result of failure to recognize the impact of development on the glucuronidation of chloramphenicol and its implications to age related individualization of therapy. The gray baby syndrome occurred in premature and newborn infants receiving high or unmodified doses of chloramphenicol and this condition can be avoided by reduction of dosage and by monitoring levels of drug in the serum of these infant [39]. Furthermore, mutation of the promoter region of UGT 1 gene has been associated with Gilbert's syndrome, a milder form of congenital unconjugated hyperbilirubinemia [40]. Sulfo-transferase (SULT1A3) activity is absent in human liver, but expressed at high levels early in fetal development, and decreases significantly in the late fetal and early neonatal development [41].

Changes in phase II drug-metabolizing enzyme expression during development, as well as the balance between phase I and phase II enzymes, can significantly alter the pharmacokinetics for a given drug or toxicant. Understanding the ontogeny of drug-metabolizing enzymes in the neonate is very important for defining the dosage regimens suitable for children and for limiting the risk of accumulation leading to adverse effects and toxicity.

1.3 Transporter Influences During Pregnancy

The drug transporters are another significant determinant in drug bioavailability and exposure. The first transporter identified was P-glycoprotein (Pgp) in 1976 [42]. Since then about 25 different transporters have been identified. The transporters can be divided into three classes. Two classes are considered uptake transporters, the SLC or solute-linked carrier transporter family and the SLCO or solute-linked carrier organic anion transporter family. The third class, the efflux transporter family, is denoted as ABC or ATP-binding cassette transporter superfamily. Notable members of this efflux family are Pgp, the multidrug-resistant proteins (MDR), the multidrug resistance associated proteins (MRPs) and the breast cancer resistant protein (BCRP) [43]. The distribution of the transporters, representative substrates, inhibitors, and inducers are also given. As shown by Shugarts and Benet [44] the intestine expresses several transporters controlling the uptake such as MCT1 (monocarboxylate transporter protein), PEPT1 and 2 (peptide transport protein), OATP 1A2 and 2B1 (organic anion transporting protein), OCT3 (organic cation protein), and others. There are several efflux transporters including Pgp, several MRPs, BCRP, MCT1, and ENT 1 and 2 (equilibrative nucleoside transporter) proteins. The liver also expresses several uptake and efflux

transporters. The hepatic uptake transporters from the blood stream include the OCTs 1 and 2, OAT2, OATPs 1B1, 1B3, 2B1 AND 1A2, NTCP (sodium-taurocholate co-transporting protein), and several MRPs 3, 4, and 5. The majority of the hepatic efflux transporters remove their compounds into the bile canaliculi, Pgp, MDR3, MRP2, BCRP, and BSEP (bile salt export pump), while one type removes compounds to the blood stream, the MRPs 3, 4, and 5 [45].

The changes in transporters in the adult female following pregnancy are not clearly understood as yet. The variety of important medications given during pregnancy such as anticancer agents, antiviral agents, and cardiovascular drugs such as warfarin can have their pharmacokinetics, their absorption, disposition, metabolism, and elimination affected in a number of ways based on the activity of the individual transporters involved or the cytochrome P450 enzymes as discussed above. While the mother's exposure and drug disposition is controlled by her own complement of cytochrome P450 enzymes, transporters, and internal hormonal and other chemical signaling systems, the drug exposure to the developing embryo and fetus is controlled primarily by the placenta and the ability of the fetus itself to handle the individual medicine given to the mother through its own complement of cytochrome P450 enzymes as discussed below.

A large number of known functional drug transporters have been found in human placenta [46]. Transporter knockout animal studies have shown the role of drug transporters in protecting the fetus from chemical effects [47]. The protection is in part due to the presence of various efflux transporters in the placenta. The effect of placental transporters in effluxing drugs such as glyburide and numerous protease inhibitors from the fetal circulation offers the potential to manipulate the passage of drugs across the placenta [48]. It is important to take into considerations, that placental transporters are vital in modulating the exposure of the fetus to drugs and, therefore, the efficacy and toxicity of such drugs towards the fetus [49]. Some of these transporters are under hormonal regulation in the placenta. Vore and Leggas [50] reported that ABCG2/BCRP expression is regulated by Estradiol and progesterone in BeWo cells, a human trophoblastic cell line.

1.4 Enzyme and Transporter Influences in the Fetus and Placenta

The placental has the ability to metabolize drugs in early pregnancy. Indeed the placenta expresses a wider variety of enzymes during the first trimester than at term [51]. Depending on the substrate, this metabolic action may have significant clinical implications on how it affects the fetus [52]. Also the developing fetus has been shown to express a number of CYP450 enzymes during its development and thus is fully capable of metabolizing endogenous and xenobiotic compounds and drugs it is exposed to. The CYP450 enzymes found to be present in the fetal liver include CYPs 1A1, 1B1,

2C8, 2D6, 2E1, 3A4, 3A5, and 3A7 after the embryonic phase (after 8–9 weeks of gestation) [23]. Xenobiotic metabolism activity was also found to be significant earlier, during organogenesis (before 8 weeks of gestation). Extra hepatic tissues such as the kidney and adrenals also contain substantial levels of CYP enzymes and can thus also exhibit metabolizing activity. The adrenals are involved in the metabolism of hormones of fetal or placental origin to help maintain and protect the fetus during gestation. The polymorphic expression of CYP3A5 and the variability of CYP3A7 expression in fetal liver were demonstrated by Hakkola et al. [53]. This suggests the existence of interindividual differences in the metabolism of xenobiotics at the prenatal stage which may contribute to individual pharmacological and/or toxicological responses in the fetus.

The placenta is also an extremely important organ for the mother and fetus. The human placenta oxidizes several xenobiotics and it represents a critical barrier from toxic agents as well as an essential organ to provide the fetus with nutrients and appropriate gas exchange during gestation. It is also active in drug metabolism and drug transport. CYP1A1, 2E1, 3A4, 3A5, 3A7, and 4B1 have been detected in the term placenta. Although little is known about phase II enzymes in the placenta, however, uridine diphosphate glucuronosyltransferases, have been detected suggesting a significant role of this enzyme in placental drug detoxification [54]. From studies in women examining the effects of smoking and found they found that placental CYP1A1 is highly inducible in pregnant women who smoke, in addition to maternal hepatic CYP1A1 and it is the most important metabolizing enzyme of the placenta for which relevant inducible activity has been demonstrated throughout pregnancy [55]. Aromatase, CYP19, and cholesterol side-chain cleaving, CYP11B genes and proteins are catalytically active in human placenta throughout the pregnancy [56].

Transport proteins play an important role in the adsorption, distribution, and elimination of a wide variety of drugs. It is therefore, comprehensible that transporter-based drug interactions can occur in the clinic. Transporter-based drug interactions in the clinic may be inhibitory, inductive, or both, and may involve influx or efflux transporters [57]. The existence of uptake and efflux transporters in organs responsible for drug biotransformation and excretion gives transporter proteins a unique gatekeeper function in controlling drug access to metabolizing enzymes and excretory pathways [44]. The presence of efflux transporters, P-glycoprotein (Pgp), the breast cancer resistance protein (BCRP), and the multidrug resistance associated proteins (MRP) in the placenta has been implicated to offer the fetus protection from medication taken during pregnancy because of their location on brush border membranes of the placenta syncytiotrophoblast [58].

Transporters for 5-HT (SERT) and NE (NET) are also expressed at the apical surface of the placenta and regulate

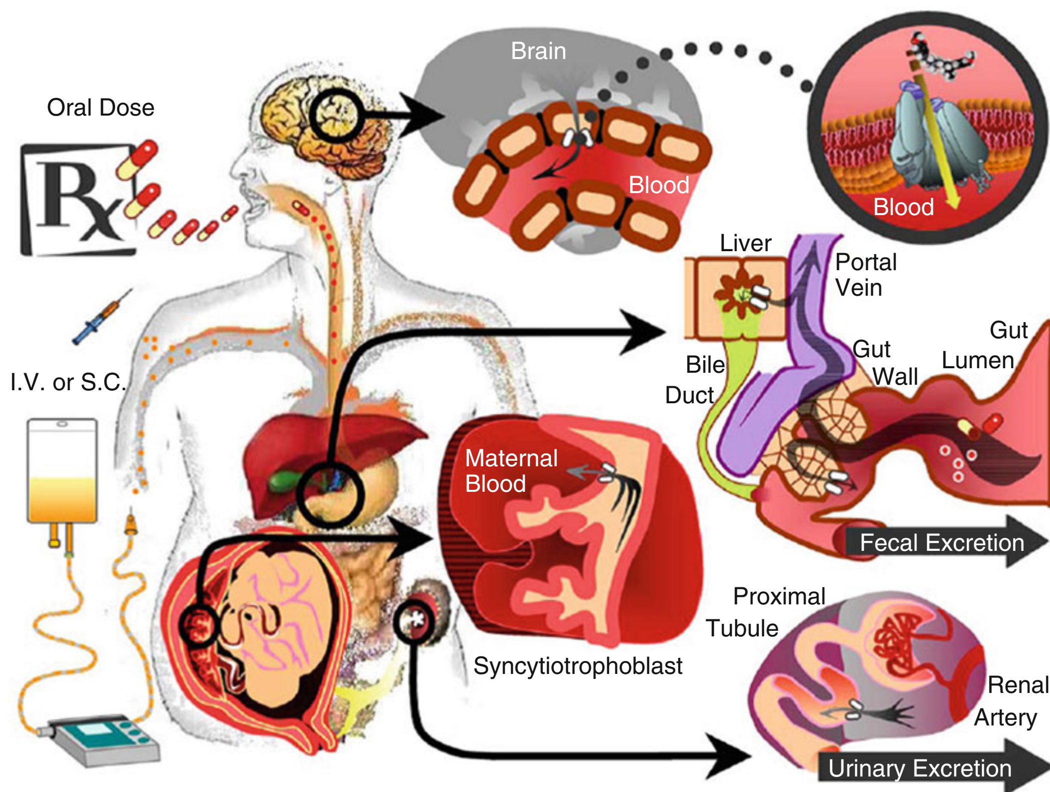

Fig. 1 In humans, P-gp is present in several tissues important for drug absorption, distribution, and elimination, such as the apical membrane of intestinal epithelial cells, the canalicular membrane of the hepatocytes, the capillary endothelial cells of the brain, the apical membrane of the placental syncytiotrophoblasts, and the apical membrane of the renal proximal tubular cells. In these tissues, P-gp functions as an efflux pump, preventing the entry of xenobiotics into these tissues (from ref. [57])

extracellular concentrations of monoamines. Some of the members of the organic anion transporters are also expressed at the basolateral surface of the syncytiotrophoblast [59]. The expression profile of these transporters varies with advancing gestation. P-gp has been shown to decline near term, leaving the fetus susceptible to potentially developmental toxic drugs commonly administered to pregnant women [60].

In the placenta, P-gp is located on the maternal-facing membrane of the syncytiotrophoblasts (Fig. 1) [58], and has been shown to play a significant role in protecting the fetus from xenobiotics [49]. However, studies in pregnant Mdr1a/b (+/+) mice, produced increase in fetal drug distribution following oral administration of the Pgp inhibitors, PSC833 or GF120918, thus indicating that the Pgp protective barrier can be ablated through pharmacological means [61].

These proteins are members of the ABC or ATP-binding cassette transporter superfamily [62]. Solute carrier (SLC) and

ATP-binding cassette (ABC) transporters play also pivotal roles in the transport of both nutrients and drugs into breast milk, thus drug–nutrient transport interactions at the lactating mammary gland are possible [63].

While most have been found to have mainly physiological substrates there are a number of drugs that also gain access to the fetus through transport across the placenta. As discussed by Hodge and Tracy [64] due to changes in many physiological parameters, the variability in the activity of the maternal drug-metabolizing enzymes as well as the influence of the drug transporters in the placenta, the exposure, efficacy, and toxicity of many drugs used by pregnant women can be difficult to predict. Transporters play an important role in exposure of the embryo/fetus to drugs with teratogenic potential during pregnancy, although the significance of placental transporters on human fetal drug exposure is almost an unstudied field so far.

1.5 Resulting Implications in Drug Disposition (DMPK) During Pregnancy

The pregnant woman presents many changes for proper drug administration as discussed above. The variety of physiological changes, the variety in the responses of the cytochrome P450 enzymes in terms of induction and inhibition as well as the presence of polymorphic forms which may be present and the influence of the drug transporters make predicting the pharmacokinetics and pharmacodynamics of any given medicine difficult. There is a lack of full information on these changes and influences that needs more investigation. Because experimenting on humans is limited the need for better animal models, in vitro systems and predictive software is needed. During pregnancy opposing changes in drug metabolism are reported to occur. This includes decreased activity of CYP1A2 and increased activity of CYP2D6 and CYP3A [31]. The CYP1A2, CYP2D6, and CYP3A enzymes are shown to be important in the metabolism of several drugs that are administered during pregnancy of coexisting conditions. Inhibitors of CYP1A2, which plays a role in metabolism of clozapine and olanzapine, include fluvoxamine and grape juice in large quantities; cigarette smoke is considered to be an inducer of enzymes. Inhibitors of CYP3A4 include erythromycin, carbamazepine, rifampin, and glucocorticoids. Women with epilepsy do have increased risks for maternal and fetal complications as children born to mothers taking antiepileptic drugs (AEDs) are at increased risk for findings of fetal anticonvulsant syndrome. In this situation the risks associated with drug exposure to the fetus and newborn need to be balanced against the risks incurred by seizures, and knowledge of pharmacokinetic alterations becomes particularly important for AED optimization. Pregnancy can affect the pharmacokinetics of AEDs at any level from absorption, distribution, metabolism, to elimination. The effect varies depending on the type of AED. The most pronounced decline in serum concentrations is seen for AEDs that are

eliminated by glucuronidation (UGT), in particular lamotrigine where the effect may be profound [65]. The apparent clearance of lamotrigine increases by 50–90 % in pregnancy, requiring dosage adjustment to prevent exacerbation of seizures [66].

These risks can be considerably reduced with careful selection of AED treatment regimens. Prescribing AEDs for women during their childbearing age should include the constant consideration of pregnancy, planned or unplanned [67].

Drug interactions involving antiviral agents mostly reflect shared toxicity with other agents (e.g., neutropenia with ganciclovir and zidovudine, pancreatitis with didanosine and alcohol), although renal excretion or hepatic metabolism may be implicated. Given the possibility of severe adverse reactions and drug interactions, antiviral chemotherapy should only be used for potentially serious virus infections during pregnancy [68].

Maternal ethanol consumption during pregnancy and lactation inhibits the hepatic metabolism of drugs such as chlorpromazine which require glucuronidation for their detoxification. This ethanol-mediated inhibition is largely exerted through the decrease in the NAD-dependent conversion of UDP-glucose (UDPG) to UDP-glucuronic acid, (UDPGA) [69].

In the fetus, important factors influencing drug metabolism are the variety of CYP450s that exist, some polymorphic, and some with changing activity in opposing directions thus presenting complicating situations on what to expect pharmacokinetically and pharmacodynamically relative to the situation in the nonpregnant state. The placenta is a very active and integral tissue in the fetal exposure to drugs. With the presence of both several CYP450s and drug efflux transporters, and there may be others as yet unknown, the placenta plays a very active role controlling the exposure of the fetus to drugs taken by the mother. The opinion on the implications for exposure and disposition is mixed. Depending on the drug and the enzyme and transporter involved the clinical response may be significant or uneventful. Therefore, for drugs with a narrow therapeutic window or those with marked pharmacologic or toxicological outcomes that are also cleared predominantly by a single CYP450 or handled by a single transporter, the need for systemic monitoring of plasma concentration to monitor exposure is warranted, at least during the initial days of starting a medication. In addition, improved understanding of transplacental drug transfer and metabolism will result in further enhancement of the clinical treatment of fetal diseases/conditions.

References

1. Glover DD, Amonkar M, Rybeck BF, Tracy TS (2003) Prescription, over-the-counter, and herbal medicine use in a rural, obstetric population. Am J Obstet Gynecol 188:1039–1045
2. Isoherranen N, Thummel KE (2013) Drug metabolism and transport during pregnancy: how does drug disposition change during pregnancy and what are the mechanisms that

cause such changes? Drug Metab Dispos 41 (2):256–262

3. Danielson PB (2002) The cytochrome P450 super family: biochemistry, evolution and drug metabolism in humans. Curr Drug Metab 3:561–597
4. Guengerich FP (1994) Catalytic selectivity of human cytochrome P450 enzymes: relevance to drug metabolism and toxicity. Toxicol Lett 70:133–138
5. Slaughter RL, Edwards DJ (1995) Recent advances: the cytochrome P450 enzymes. Ann Pharmacother 29:619–624
6. Mizuno N, Niwa T, Yotsumoto Y, Sugiyama Y (2003) Impact of drug transporter studies on drug discovery and development. Pharmacol Rev 55(3):425–461
7. Borst P, Evers R, Kool M, Wijnholds J (2000) A family of drug transporters: the multidrug resistance-associated proteins. J Natl Cancer Inst 92:1295–1302
8. Rogers AS (1994) The role of cytochrome P450 in developmental pharmacology. J Adolesc Health 15(8):635–640
9. Hollenberg PF (2002) Characteristics and common properties of inhibitors, inducers, and activators of CYP enzymes drug metabolism reviews. Drug Metab Rev 34(1&2):17–35
10. Feghali MN, Mattison DR (2011) Clinical therapeutics in pregnancy. J Biomed Biotechnol ID 783528, 13 p
11. Little B (1999) Pharmacokinetics during pregnancy: evidence-based maternal dose formulation. Obstet Gynecol 93:858–868
12. Frederiksen MC (2001) Physiologic changes in pregnancy and their effect on drug disposition. Semin Perinatol 5:120–123
13. Pavek P, Ceckova M, Staud F (2009) Variation of drug kinetics in pregnancy. Curr Drug Metab 10:520–529
14. Mitani GM, Steinberg I, Lien EJ, Harrison EC, Elkayam U (1987) The pharmacokinetics of antiarrhythmic agents in pregnancy and lactation. Clin Pharmacokinet 12:253–291
15. Best B, Capparelli EV (2008) Implications of gender and pregnancy for antiretroviral drug dosing. Cur Opin HIV AIDS 3:277–282
16. Morgan DJ (1997) Drug disposition in mother and fetus. Clin Exp Pharmacol Physiol 24:869–873
17. Jeong H (2010) Altered drug metabolism during pregnancy: hormonal regulation of drug-metabolizing enzymes. Expert Opin Drug Metab Toxicol 6(6):689–699
18. Loebstein R, Lalkin A, Koren G (1997) Pharmacokinetic changes during pregnancy and their clinical relevance. Clin Pharmacokinet 33:328–343
19. Krishna DR, Klotz U (1994) Extrahepatic metabolism of drugs in humans. Clin Pharmacokinet 26:144–160
20. O'Shaughnessy PJ, Monteiro A, Bhattacharya S, Fowler PA (2011) Maternal smoking and fetal sex significantly affect metabolic enzyme expression in the human fetal liver. J Clin Endocrinol Metab 96(9):2851–2860
21. Huang SM, Temple R, Throckmorton DC, Lesko LJ (2007) Drug interaction studies: study design, data analysis, and implications for dosing and labeling. Clin Pharmacol Ther 81(2):298–304
22. Sasaki S, Kondo T, Sata F, Saijo Y, Katoh S, Nakajima S, Ishizuka M, Fujita S, Kishi R (2006) Maternal smoking during pregnancy and genetic polymorphisms in the Ah receptor, CYP1A1 and GSTM1 affect infant birth size in Japanese subjects. Basic Sci Reprod Med 12(2):77–83
23. Blake MJ, Gaedigk A, Pearce RE, Bomgaars LR, Christensen ML, Stowe C, James LP, Wilson JT, Kearns GL, Leeder JS (2005) Ontogeny of dextromethorphan *O*- and *N*-demethylation in the first year of life. Clin Pharmacol Ther 81:510–516
24. Treluyer JM, Jacqz-Aigrain E, Alvarez F, Cresteil T (1991) Expression of CYP2D6 in developing human liver. Eur J Biochem 202:583–588
25. Fakhoury M, Jacqz-Aigrain E (2005) Developmental pharmacogenetics. Pediatrica 16 (2):28–31
26. Stevens JC, Marsh SA, Zaya MJ, Regina KJ, Divakaran K, Le M, Hines RN (2008) Developmental changes in human liver CYP2D6 expression. Am Soc Pharmacol ExpTher 36 (8):1587–1593
27. Zanger UM, Turpeinen M, Klein K, Schwab M (2008) Functional pharmacogenetics/genomics of human cytochromes P450 involved in drug biotransformation. Anal Bioanal Chem 39(6):1093–1108
28. De Wildt SN, Kearns GL, Hop WC, Murry DJ, Abdel-Rahman SM, van den Anker JN (2001) Pharmacokinetics and metabolism of intravenous midazolam in preterm infants. Clin Pharmacol Ther 70:525–531
29. Davis M, Simmons CJ, Dordoni B et al (1973) Induction of hepatic enzymes during normal human pregnancy. J Obstet Gynaecol Br Commonw 80:690–694
30. Wadelius M, Darj E, Frenne G, Rane A (1997) Induction of CYP2D6 in pregnancy. Clin Pharmacol Ther 62:400–407

31. Tracy TS, Venkataramanan R, Glover DD, Caritis SN (2005) Temporal changes in drug metabolism (CYP1A2, CYP2D6 and CYP3A activity) during pregnancy. Am J Obstet Gynecol 192:633–639
32. Koh KH, Xie H, Yu A-M, Leong H (2011) Altered cytochrome P450 expression in mice during pregnancy. Drug Metab Dispos 39 (2):165–169
33. Strickler S, Margaret A, Miller E, Linda A, Marie-Helene DS, Spielberg SP (1985) Genetic predisposition to phenytoin-induced birth defects. Lancet 2:746–749
34. Shanks MJ, Wiley MJ, Kubow S, Wells PG (1989) Phenytoin embryotoxicity: role of enzymatic bioactivation in a murine embryo culture model. Teratology 40:311–320
35. Wilffert B, Altena J, Tijink L, van Gelder M, de Jong-van den Berg L (2011) Pharmacogenetics of drug-induced birth defects: what is known so far? Pharmacogenomics 12(4):547–558
36. Nagai G, Ono S, Yasui-Furukori N, Nakamura A, Mihara K, Kondo T (2009) Formulations of valproate alter valproate metabolism: a single oral dose kinetic study. Ther Drug Monit 31 (5):592–596
37. Shaffer CL, Gal P, Ransom JL, Carlos RQ, Smith MS, Davey AM, Dimaguila MA, Brown YL, Schall SA (2002) Effect of age and birth weight on indomethacin pharmacodynamics in neonates treated for patent ductus arteriosus. Crit Care Med 30:343–348
38. McCarver DG, Hines RN (2002) The ontogeny of human drug-metabolizing enzymes: phase II conjugation enzymes and regulatory mechanisms. J Pharmacol Exp Ther 300 (2):361–366
39. Feder HM, Osier C, Maderazo EG (1981) Chloramphenicol: a review of its use in clinical practice. Clin Infect Dis 3(3):479–491
40. Bosma PJ, Chowdhury JR, Bakker C, Gantla S, de Boer A, Oostra BA, Lindhout D, Tytgat GN, Jansen PL, Oude Elferink RP, Chowndry NR (1995) The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. N Engl J Med 333 (18):1171–1175
41. Richard K, Hume R, Kaptein E, Visser TJ, Coughtrie MWH (2001) Sulfation of thyroid hormone and dopamine during human development: ontogeny of phenol sulfotransferases and arylsulfatase in liver, lung and brain. J Clin Endocrinol Metab 86:2734–2742
42. Juliano RI, Ling VA (1976) A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochim Biophys Acta 455:152–162
43. Klaassen CD, Aleksunes LM (2010) Xenobiotic, bile acid, and cholesterol transporters: function and regulation. Pharmacol Rev 62 (1):1–96
44. Shugarts S, Benet LZ (2009) The role of transporters in the pharmacokinetics of orally administered. Pharm Res 26(9):2039–2054
45. Scherrmann JM (2009) Transporters in absorption, distribution, and elimination. Chem Biodivers 6(11):1933–1942
46. Vähäkangas K, Myllynen P (2009) Drug transporters in the human blood-placental barrier. Br J Pharmacol 158(3):665–678
47. Klaassen CD, Lu H (2008) Xenobiotic transporters: ascribing function from gene knockout and mutation studies. Toxicol Sci 101:186–196
48. Gedeon C, Koren G (2006) Designing pregnancy centered medications: drugs which do not cross the human placenta. Placenta 27 (8):861–868
49. Unadkat JD, Dahlin A, Vijay S (2004) Placental drug transporters. Curr Drug Metab 5 (1):125–131
50. Vore M, Leggas M (2008) Progesterone acts via progesterone receptors A and B to regulate breast cancer resistance protein expression. Mol Pharmacol 73:613–615
51. Hakkola J, Pelkonen O, Pasanen M, Raunio H (1996) Xenobiotic-metabolizing cytochrome P450 enzymes in the human feto-placental unit: role in intrauterine toxicity. Crit Rev Toxicol 28:35–72
52. Weier N, He SM, Li XT, Wang LL, Zhou SF (2008) Placental drug disposition and its clinical implications. Curr Drug Metab 9:106–121
53. Hakkola J, Raunio H, Purkunen R, Saarikoski S, Vähäkangas K, Pelkonen O, Edwards RJ, Boobis AR, Pasanen M (2001) Cytochrome P450 3A expression in the human fetal liver: evidence that CYP3A5 is expressed in only a limited number of fetal livers. Biol Neonate 80 (3):193–201
54. Syme MR, Paxton JW, Keelan JA (2004) Drug transfer and metabolism by the human placenta. Clin Pharmacokinet 43(8):487–514
55. Stejskalova L, Pavek P (2011) The function of cytochrome P450 1A1 enzyme (CYP1A1) and aryl hydrocarbon receptor (AhR) in the placenta. Curr Pharm Biotechnol 12(5):715–730
56. Pasanen M (1999) The expression and regulation of drug metabolism in human placenta. Adv Drug Deliv Rev 38(3):81–97
57. Endres CJ, Hsiao P, Chung FS, Unadkat JD (2006) The role of transporters in drug interactions. Eur J Pharm Sci 27(5):501–517

58. Atkinson DE, Brice-Bennett S, D'Souza SW (2007) Antiepileptic medication during pregnancy: does fetal genotype affect outcome? Pediatr Res 62(2):120–127
59. Eshkoli T, Sheiner E, Ben-Zvi Z, Holcberg G (2011) Drug transport across the placenta. Curr Pharm Biotechnol 12(5):707–714
60. Iqbal M, Audette MC, Petropoulos S, Gibb W, Matthews SG (2012) Placental drug transporters and their role in fetal protection. Placenta 33(3):137–142
61. Smit JW, Huisman MT, van Tellingen O, Wiltshire HR, Schinkel AH (1999) Absence or pharmacological blocking of placental P-glycoprotein profoundly increases fetal drug exposure. J Clin Invest 104(10):1441–1447
62. Dean M, Rzhetsky A, Allikmets R (2001) The human ATP-binding cassette (ABC) transporter superfamily. Genome Res 11(7):1156–1166
63. Gilchrist SE, Alcorn J (2010) Lactation stage-dependent expression of transporters in rat whole mammary gland and primary mammary epithelial organoids. Fundam Clin Pharmacol 24(2):205–214
64. Hodge LS, Tracy TS (2007) Alterations in drug disposition during pregnancy: implications for drug therapy. Expert Opin Drug Metab Toxicol 3:557–571
65. Tomson T, Landmark CJ, Battino D (2013) Antiepileptic drug treatment in pregnancy: changes in drug disposition and their clinical implications. Epilepsia 54(3):405–414
66. Pennell PB, Peng L, Newport DJ, Ritchie JC, Koganti A, Holley DK, Newman M, Stowe ZN (2008) Lamotrigine in pregnancy: clearance, therapeutic drug monitoring, and seizure frequency. Neurology 70:2130–2136
67. Pennell P (2005) Using current evidence in selecting antiepileptic drugs for use during pregnancy. Epilepsy Curr 5(2):45–51
68. Morris DJ (1994) Adverse effects and drug interactions of clinical importance with antiviral drugs. Drug Saf 10(4):281–291
69. Rawat AK (1981) Ethanol and psychotropic drug interaction during pregnancy and lactation. Biochem Pharmacol 30(17): 2457–2460

Methods in Pharmacology and Toxicology (2017) 17–42
DOI 10.1007/7653_2015_55
© Springer Science+Business Media New York 2015
Published online: 13 August 2016

Critical Periods of Development in Teratology

Padmanahban Rengasamy

Abstract

Several years ago Wilson (1973) postulated his principles of teratogenesis, which have been revisited previously. However, there is a strong need to evaluate critically and summarize recent advances in the area of susceptible periods in teratology. This review focused on three model teratogens, namely ionizing radiation, congenital rubella, and retinoids. Continued presence of rubella virus in the placental and fetal tissues suggests an extended period of susceptibility to abnormalities of lens and otocyst differentiation beyond the narrow window of embryonic period. The currently understood critical period for maternal radiation exposure-related microcephaly in the offspring has been derived from Hiroshima and Nagasaki experience. Data on pregnancy complications of radiation-exposed childhood cancer survivors is alarming. Cumulative use of diagnostic CT imaging and the associated ionizing radiation burden are reported to have a lifetime attributable cancer risk. This coupled with experimental data on radiation effect on neurogenesis in the cerebellum and that on global gliogenesis in the CNS is a reason to revisit the critical period definition for microcephaly in the offspring of mothers exposed to radiation. Both retinoic acid and valproic acid studies in experimental animals have demonstrated unequivocal evidence for the existence of more than one susceptible period for skeletal defects and cleft palate. It is important to realize the distinction between the "critical period for a malformation" and the "critical period for an organ as a whole." Vascular disruption and post-closure neural tube defects have been observed to occur far beyond the embryonic period of development. It is now clear that the textbook description of the embryonic period as the window of susceptible period is rift with problems. In the light of newer advances in teratology, a second look at our definition of teratogenic manifestations and sophistication of methods of ascertainment of abnormal development will help us develop effective preventative strategies in the field of birth defects.

Keywords: Teratology, Critical periods, Ionizing radiation, Congenital rubella, Retinoids, Microcephaly, Cataract, Therapeutic radiation, Childhood radiation, Neurogenesis, Gliogenesis, Vascular disruption, Post-closure neural tube defects

1 Introduction

According to CDC (Center for Disease Control, Atlanta) congenital malformations are "common, costly and critical" [1]. About 1 in 33 births in the USA is affected by birth defects which amount to 120,000 babies. The European Surveillance of Congenital Anomalies [2] estimates the prevalence rate of all congenital malformations in Europe, excluding chromosomal anomalies as 219.7 per 10,000 births (2.2 %) for 2003–2007. In certain parts of Africa where antenatal care is unavailable the prevalence rate is reported to be as high as 14 % [3]. Congenital malformations and chromosomal

anomalies are of serious public health issues because they cause infant mortality, childhood morbidity, and long-term disability in survivors. They are costly. They are a burden on the economy, families, and community as well as a leading cause of years of potential life lost [4]. Congenital malformations, such as neural tube defects (NTD), are not only highly prevalent, particularly in lower and middle income countries, but they also cost significant amounts of money to the communities that are already economically strained. Yi et al. [5] reviewed studies on economic burden of NTD covering the period 1976–2010, evaluated the direct and indirect costs of care required by affected individuals, and concluded that the benefits of prevention of NTD with maternal folic acid supplementation far outweighed the cost of care required by NTD patients. A recent study looked at 37 previous studies on the burden of NTD among middle and lower income countries and analyzed data on live births, still births, and terminations and estimated that a total of 190,000 neonates were added each year to the NTD burden in those countries [6]. The cost of raising a child with spina bifida from birth to the age of 18 years is estimated to be about $120,000 in the USA. Even greater amounts of money are spent in caring children with open spina bifida in developed countries in Europe [7]. Although not all NTD could be prevented, folate fortification and periconceptional folate supplementation have significantly reduced the incidence of NTD and improved cost saving in developing countries [8]. About a third to one half of all pregnancies are estimated to be unintended [9, 10]. It is not possible to assume that this section of women are on supplemental folate or would take care not to consume alcohol or refrain from smoking. Intended or unintended pregnancies are costly. Pregnancies complicated by situations requiring surgical procedures, preterm births, multiple births, intrauterine growth restriction (IUGR), etc. cost significantly more. Now that the beneficial effects of folate is known, there is however a concern that the probability of vitamin B12 deficiency might go unnoticed due to the masking effects of folate on B12 deficiency-related anemia and the consequent neurological damage to the patients [11]. Maternal smoking during pregnancy is known to cause orofacial clefts affecting about 6 % of births. If smoking during pregnancy is prevented, it can result in an estimated saving of $40.4 million for each annual cohort of 430 averted cases [12]. Other pregnancy outcomes that are of interest to teratologists include preterm births and small for gestational age (SGA) births. Complications associated with preterm births are attributed to be responsible for 35 % of the world's 3.1 million neonatal deaths estimated annually [13]. Preterm birth rates are increasing in economically affluent nations. Both SGA and preterm births account not only for significant proportions of perinatal mortality and morbidity but they are also a major challenge for maternal and perinatal care globally [14]. Preterm birth rates are

reported to be increasing in developed countries [15]. Teratologists should also be equally concerned about the high prevalence rates of intrauterine growth retardation (IUGR) which is often complicated by accompanying congenital malformations. Growth restriction may be primary predisposing the fetus to malformations or secondary to the malformation or share a common etiology with the coexisting physical anomaly [16]. There is a strong body of evidence supporting the association between stillbirth, IUGR, and congenital malformations. In fact IUGR is clinically employed for the stillbirth prediction. Stillbirth and IUGR are reported to have overlapping etiology and risk factors [17, 18]. Individuals who are growth restricted in utero are now known to develop obesity, type 2 diabetes, and/or cardiovascular diseases later during adult life possibly due to hypomethylation of DNA and other epigenetic modifications [19–22].

To eliminate or minimize the huge economic burden of congenital malformations, prematurity, and IUGR to the affected individuals, their families, and the health care system, efforts need to be directed at prevention. Planning preventative strategies in dealing with the issue of congenital malformations, small for gestational age (SGA), and IUGR requires a thorough understanding of the normal processes of development and the causes, molecular, cellular, and morphological mechanisms and both anatomical and functional manifestations of abnormal processes of development. Historically speaking, people have responded to the birth of a malformed baby with fear, awe, and deep concern. The literature on the history of congenital malformations includes descriptive accounts often stemming from mythology with etiology attributed to maternal impressions during pregnancy, the position of stars, divine retribution for wrong doing, etc. However, concurrently there has always been some form of scientific inquiry attributing congenital anomalies to various causes including fetal distress, arrest of embryonic growth and interference with development by chemical substances, and mechanical factors.

William Harvey (1561) (cited by Wilson [23]) first advanced his theory of developmental arrest meaning that developmental anomalies observed in newborns represented the appearance of the organ primordia at the time point when development was interfered with by an environmental agent. Etienne Geoffroy Saint-Hilaire (1772–1844) and his son Isidore Geoffroy Saint-Hilaire (1805–1861) studied embryology, comparative anatomy, and experimental teratology in an effort to explain congenital malformations. They designed experiments on chick embryos to alter development by various physical manipulations including high and low temperatures. This was possibly the first recorded scientific inquiry into environmental effects on embryonic development. The results of their studies helped shift the focus from mythology to an environmental etiology of abnormal embryonic development and thus teratology became a "special branch of the greater science

of organization" [24–26]. This work was extended by Camille Dareste's experiments in which developmental stage-dependent monstrosities were induced in chick embryos by mechanical stimulation with heat, cold, shock, etc. during incubation [27]. Dareste believed that his experimental manipulations somehow slowed down or caused arrest of the normal developmental processes. Dareste's work appears to have deeply impacted Stockard [28] who experimented with the common minnow (*Fundulus heteroclitus*) embryos by subjecting them to variable temperatures or variable concentrations of O_2 and CO_2, thus causing development of the embryos to slow down (by moderately low temperature, for example) or to stop (by very low temperature, for example) for given periods of time after which restoring them to the normal environment. Depending on the developmental stage at which the environment of the embryos was first altered, embryos with too slow rates of resumption were found to result in significant malformations while stoppage was followed by high mortality. Some periods ("moments") of development when by experimental manipulation developmental arrest or interruption could be caused without embryonic death or malformations were classed as ***indifferent moment***. "***Moments of supremacy***" were when primordia of organs were forming; these were developmental stages which could not be arrested without causing serious effects on subsequent development. Such sensitive periods constituted "***critical moments***" for Stockard. He claimed that all types of malformations not of hereditary origin were simply the result of developmental arrest. This belief led him to the following general principles ("or propositions") of abnormal development (quoted from Stockard [28]):

- "All types of monsters may be caused by the one and the same experimental treatment
- Any one type of monster, such as cyclopia may be produced by a great number of different experimental treatments
- All effective treatments tend primarily to lower the rate of development
- The type of monster induced depends upon the particular developmental moment or moments during which the developmental rate was reduced"

Stockard's propositions and other theories of birth defects, such as "teratogenic termination period" (of Schwalbe) and "critical moments" (periods) of Stockard [28] were possibly deeply influenced by Dareste's postulates of teratogenesis, which are (quoted from Jelineck [24, 28]) as follows:

- "Identical defects can be induced by administering different agents
- Particular embryos react to an adverse condition in dissimilar ways

- The dissimilarity is caused by unequal combinations of inherited gifts (genes) and extrinsic influences
- Type of defect depends upon the strength and time of action of adverse impulse
- The smaller the defect, the later it will become apparent"

For these reasons Dareste is regarded as the founder of experimental teratology, although Dareste himself bestowed that distinction and honor upon Etienne St Hilaire. There is a short but excellent summary account of the history of teratology in the first half of 1900 [29].

Although Mendel postulated his theory of inheritance in as early as 1865, it was only in 1900 his postulates became widely known when Hugo de Vries, Carl Correns, and Erich von Tschermak-Seyenegg also arrived at the same conclusions independently and unaware of Mendel's work [30]. Soon the medical community started easily attributing almost all birth defects to a genetic cause. Subsequently the discovery that maternal exposure to ionizing radiation, maternal rubella infection, and maternal dietary excess and deficiency of micronutrients during pregnancy could cause congenital anomalies in the offspring convinced the scientific community beyond doubt that malformations might arise as a result of several non-genetic causes. The situation began to change rapidly and extensively when Warkany and his group published results of a series of experiments they conducted in laboratory animals showing that maternal nutritional deficiencies could induce predictable frequencies of fetal malformations. The thalidomide embryopathy reported in 1960s [31, 32] was an unexpected catastrophe which took the teratologists around the world by surprise. It brought to light the possibility of "innocent" prescription drugs also being highly teratogenic to human embryos. It led to rapid expansion and greater scope of the field of teratology and reproductive toxicology. Some knowledge of prenatal development is important in understanding the concept of embryonic susceptibility to develop malformations.

2 Developmental Periods

Clinically pregnancy is commonly divided into three sequential trimesters, each lasting 3 months. Developmentally the transition from one trimester to the next is not abrupt but a continuum. Some embryologist would divide development into three periods, namely the period of implantation, embryonic period, and fetal period, though clinicians who also pay greater attention to developmental-stage-specific morphogenetic processes would divide development into (1) blastogenesis—up to 2 weeks postconception, (2)

organogenesis—end of week 2 (or beginning of week 3) to end of week 8 postconception, (3) fetogenesis—early (9–15 weeks), mid (16–25 weeks), and late (26–38 weeks) fetal periods [33]. The highlights of the first trimester include events such as implantation, differentiation of the inner cell mass into three germ layers, craniocaudal and lateral foldings of the embryo, embryonic body formation, organogenesis, fetal and placental membranes development, and establishment of a functional circulatory system. During the second and third trimesters, the organ primordia continue to grow and elaborate and differentiate in complex ways so that the embryonic organ systems attain a certain level of functional maturity. The fetus as a whole grows in size and with time attains definitive body proportions. Keeping pace with the fetus, the placenta grows in size, differentiates further, and secretes a myriad of substances that are of major metabolic and developmental significance. The placenta plays an active role in materno-fetal exchange processes and supports fetal growth. Although short lived, in terms of its structural complexity and functional diversity the placenta is possibly the most versatile of all organs. Intrauterine development and growth are tightly controlled both spatially and temporally by a series of morphoregulatory processes but remain subject to epigenetic modifications [34, 35]. Developmental control mechanisms are not invincible to environmental influences. The sensitivity of the embryo to the modifiers appears to differ at different time points during development and is species specific [36, 37].

3 Principles of Teratogenesis

Based on his own observations in his experimental studies and critical evaluation of the literature of his time Jim Wilson, one of the founding fathers of Teratology Society, first formulated his five Principles of Teratology in 1959 [38], which he subsequently revised and restated in his monograph *Environment and Birth Defects* in 1973 [23] and in other publications. For the purpose of understanding of our readers, Wilson's Principles of Teratology [23] are stated here:

1. "Susceptibility to teratogenesis depends on the genotype of the conceptus and the manner in which this interacts with adverse environmental factors
2. Susceptibility to teratogenesis varies with developmental stage at the time of exposure to an adverse influence
3. Teratogenic agents act in specific ways on developing cells and tissues to initiate sequences of abnormal developmental events (pathogenesis)
4. The access of adverse influences to developing tissues depends on the nature of the influence

5. The four manifestations of deviant development are embryonic death, malformations, intrauterine growth restriction and functional deficit
6. Manifestations of deviant development increase in frequency and degree as dosage increases from the no effect to the totally lethal level"

Since Wilson (1973) enunciated the Principles of Teratology [23], the field of Teratology has had significant advances in keeping pace with the progress in genetics, molecular biology, and developmental biology and we have come to realize how the recent advances in the realm of teratogenic mechanisms have only validated and strengthened further the scientific bases of Wilson's Principles of Teratology. This review focuses on developmental periods of embryonic susceptibility to exposure to adverse influences.

Charles Stockard (C R Stockard 1879–1936) made some milestone contribution to developmental biology, aging, alcohol effect on embryonic development, etc. However, his most outstanding contribution to teratology is the concept of "critical moments" which is based on the results of his experiments with fish embryos [28, 39]. Although many of us teratologists who belong to the post Warkany-Wilson era do not agree with all that Stockard published, there is no denial that Stockard is a milestone in the ontogenesis of the "Principles of Teratology" and that he will continue to be cited in teratology literature. At a time when there weren't electron microscopy and PCR, simply by subjecting the eggs of minnows and trout to cold temperatures Stockard [28] produced a variety of cyclopic and two headed embryos and observed that the window of opportunity to produce these anomalies was limited to the first 24 h after fertilization of the egg. If the exposure was delayed beyond 24 h nothing remarkable happened. He called these moments "moment of supremacy" and "moment of indifference," respectively, which later became referred to as "sensitive/critical period" and "refractory period," respectively. The concept of "critical moments" now widely referred to as "critical periods" and Schwalbe's theory of "teratogenic termination periods" possibly inspired Wilson [23], thus leading to his postulate #2 "Susceptibility to teratogenesis varies with the developmental stage at the time of exposure to an adverse influence."

4 Validating the Principle of Critical Period

It is outside the scope of this review to discuss critical periods in relation to all teratogens. It will discuss critical periods in relation to *a physical teratogenic agent*—ionizing radiation, *an infectious teratogenic agent*, rubella, and *a chemical teratogenic agent*—the retinoids.

4.1 Ionizing Radiation

Ionizing radiation is possibly the earliest and extensively researched example of human teratogens [40, 41] thought to be in support of the concept of Wilson's postulate on developmental stage-dependent susceptibility to teratogenesis of human embryos. Radiation-affected pregnancies of the Nagasaki-Hiroshima atomic bomb explosion episodes highlighted for the first time the deleterious maternal and fetal complications of large doses of radiation. Congenital malformations and fetal growth restriction observed subsequently in numerous pregnancies exposed to radiation for diagnostic and therapeutic reasons were described to be development stage dependent [42–44]. Humans are exposed rather continuously to low doses of radiation and at low dose rates. The data of vast majority of reports published over the past 100 years on reproductive outcome of radiation-exposed human pregnancies come from Hiroshima and Nagasaki involving high doses emanating directly from atomic bomb explosions. Otake et al. [45] reviewed this literature and suggested 8–15 weeks of post-fertilization period as the most sensitive period for severe mental retardation in the offspring of women exposed to radiation when they were pregnant. Several investigators seem to tacitly approve a low-dose radiation-sensitive period of 8–15 weeks after conception too [33, 46]; however, they also suggest that accumulation of radiation effects via repeated-diagnostic procedures in a pregnant woman could result in microcephaly and severe mental retardation in the offspring. Epidemiologic and laboratory animal studies with an acute dose of 1 Gy [47, 48] suggest an "all or none effect" during preimplantation period, malformations, growth restriction, and microcephaly for organogenesis period and only IUGR and microcephaly for fetal period. Epidemiological studies commonly report microcephaly and mental retardation occurring in children following maternal exposure to radiation. The critical periods referred to in these studies vary extremely (gestation weeks 4–19; 8–15; 16–25, etc.; see for references De Saints et al.[47]), primarily because the doses and frequency of radiation exposures were neither consistent nor comparable across studies possibly dictated by individual patient needs. Surveying volumes of literature as well as 27 cases of pregnancies irradiated for pelvic diseases, Dekaban [44] drew a tentative timetable of abnormalities showing the period of 3.5–20.0 weeks of gestation as the critical period of radiation-induced fetal malformations and IUGR. He noticed that with advancing gestational age, there was a progressive decline in the incidence of irradiation-related malformations and reduction in severity of IUGR. Apparently this timetable is in sharp contrast to the developmental stage- and dose-dependent organ system-specific malformations he observed in a mouse study he did with 200 R dose of irradiation administered on each consecutive day between 7th and 18th days of gestation. The data from this study are similar in many respects to those of numerous experimental

studies reported in the literature. A common inference from these experimental animal studies and human clinical studies is a warning that the application of the concept of critical period for radiation-induced malformations is rift with problems and pitfalls. In the development of the nervous system, neurogenesis of the cerebellum and dentate gyrus of the cerebrum as well as global gliogenesis in CNS in particular are events that are largely late gestational and early postnatal in timing; these observations are reinforced by robust data from experimental studies that substantiate the fact that proliferative and migratory cell populations are vulnerable to environmental and nutritional interferences. Robust experimental data are available showing that low-dose radiation of late gestational and young postnatal animals gives rise to CNS morphological and behavioral abnormalities [49–53] and therefore careful interpretation of the structural and functional consequences of irradiation at different time points in life of the organism and our definition of what constitutes a teratogenic outcome (spontaneous abortion or resorption in lab animals, gross anatomical malformation, growth restriction, or functional disorders) [54]) will help us define "sensitive period/critical period or teratogenic termination period vs. period of organogenesis." Any conclusions on critical period for radiation would only make sense if one takes into account the dosage—acute vs. chronic exposure, maternal and fetal genetic makeup, concurrent exposure to prescription and nonprescription therapeutic agents, maternal nutritional status, etc. Also important to realize here is the distinction between the "critical period for a malformation" and the "critical period for the organ as a whole" [54]. In the case of brain open defects of the neural tube such as anencephaly might occur over a narrow window of time during neural tube closure, whereas neuronal/glial cell deficits might result from influences that interfere with cell proliferation, differentiation, and/or migration that characterize longer windows of growth spurts later in development [55]. Defining such differential susceptibility in humans is complex. Recognition and appreciation of this critical vulnerability as an essential aspect of development will benefit significantly if data from well-controlled animal experiments are taken advantage of for comparison with human development as described by Clancy et al. [56, 57].

One another body of data that has not received much attention of the teratologists is the time- and activity-dependent metabolic/functional maturation of the different regions of the brain. Studies of cerebral glucose utilization with positron emission tomography (PET) have revealed the existence of progressive metabolic maturation process that starts at 2–3 months of age and persists until ages of 16–18 years. The periods of neuronal and glial proliferation, progressive synaptogenesis, production of neurotransmitters, and various receptors are followed by postnatal periods of selective apoptosis. Surviving neurons also go through a pruning phase

characterized by selective synaptic elimination and repeated activity-dependent stabilization of certain circuits, thus molding an optimum cytoarchitecture of the CNS appropriate for the future needs of the individual [58, 59]. Patients subjected to radiologic evaluations or treated with radiation for neoplastic diseases become exposed to amounts of radiation that are commonly assumed to be safe [60]. Computer tomography (CT) scans, by virtue of their ease of acquisition of serial multiple images in quick succession and high levels of diagnostic capabilities, have revolutionized medical practice throughout the world. However, a small percentage of obstetric and gynecologic physicians have always suspected that abdominal scans use unrealistically high levels of radiation that might harm pregnancy. For example, a chest CT scan uses about 100 times greater amount of ionizing radiation than a lateral or AP chest X-ray. Although CT only accounts for about 17 % of all medical imaging, it contributes to about 50 % of clinical imaging-derived radiation burden [61]. High-speed multiphase image acquisition results in increasing amounts of ionizing radiation exposure. Abdominal and pelvic CTs outnumber CTs of other regions of the body. More recently there has been an alarming numerical increase in the use of CT in patients of age 1–18 with a lifetime attributable cancer risk [62]. Increased risks of infertility, miscarriage, stillbirth, intrauterine growth restriction, and preterm birth have been reported in studies that looked at pregnancies of radiation-exposed childhood cancer survivors [63, 64]. These deleterious later effects of childhood exposure to radiation strongly indicate the need for further modifications of the concept of susceptible periods in teratology to include reproductive outcomes of childhood and pre-pregnancy exposures.

4.2 Maternal Rubella Infection

Maternal infections during pregnancy are attributed to be causally related to about 3 % of total malformations observed during the first year of life [65]. This figure has to be cautiously interpreted in view of the fact that rate of intrauterine infections during pregnancy is estimated to be as high as about 14 % when appropriate laboratory investigations are used for diagnosis [66] and that syphilis and rubella infections are associated with a high incidence of congenital anomalies in the offspring; in fact about 64 % women report to have had one or more infections during their pregnancies [67]; some congenital abnormalities, such as deafness and eye defects due to maternal infections, are not detected until the second year of life or even later [68]. Two German physicians (de Bergean and Orlow) are credited to be the first to describe the clinical manifestations of rubella [69]. Although popularly known as German measles based possibly on this German connection, there was some confusion about the relationship of rubella to measles and to scarlet fever. It was yet another German physician, George de Maton, who for the first time characterized rubella in 1814 as a distinct disease

entity, called "rotheln" [69–71]. Henry Veale, a British Army surgeon, is said to have named the disease rubella (from the Latin for "reddish things," the rashes) in 1866. He thought the original word "*rotheln*" sounded "harsh and foreign to our ears."

Rubella virus (RV), a member of the family *Togaviridae* and the only member of the genus *Rubivirus*, is transmitted via respiratory aerosols. The nasopharyngeal mucosa of the upper respiratory tract and the underlying lymphoid tissue are the initial sites where RV replication occurs. The upper deep cervical regional lymph nodes subsequently become involved. Although initially mild, and often asymptomatic, macropapular rashes appear within 2–3 weeks after exposure. The rashes appear first on the face and spread rapidly to the trunk and limbs. Patients develop mild fever, pharyngitis, lymphadenopathy, arthritis, and arthralgia with non-immunized women being affected more frequently than men. More serious complications such as progressive rubella panencephalitis have also been reported in some of the RV-infected patients.

Historically there was a worldwide epidemic rubella infection in 1940 and in 1941, Norman McAlister Gregg [72], an Australian ophthalmologist reported for the first time a high frequency of congenital cataract, reduced pupillary reaction to light, atrophic iris and congenital heart defects comprising patent interventricular foramen, and patent ductus arteriosus in newborn children [71, 72]. To his meticulously recorded history of 13 cases were added 65 cases of infants with cataract collected by his fellow physicians. Gregg and his fellow physicians thought of several etiological factors including the German measles that preceded the birth of numerous infants with a recognizable spectrum of congenital abnormalities, termed congenital rubella syndrome (CRS). Similar cases were reported subsequently from Japan, Europe, and the USA. Greg's report [72] showed for the first time that a virus could be a human teratogen. The rubella pandemic that occurred in the USA from 1962 through 1964 resulted in approximately 12.5 million cases of clinically established rubella, 11,000 fetal demise, 20,000 live born infants with CRS, and 2100 neonatal deaths due to congenital anomalies that constitute congenital rubella syndrome (CRS). The estimated cost to the national economy was approximately $2.0 billion [73, 74]. CRS is the cause of a preventable spectrum of malformations that comprises ophthalmic (cataract, microphthalmia, glaucoma, unusual type of nystagmus (of "course, jerky, purposeless, searching type")), cardiovascular (atrial and ventricular septal defects, widely patent ductus arteriosus, pulmonary stenosis, and coarctation of the aorta), and neurological defects (microcephaly, mental retardation, and sensory neural deafness), and intrauterine growth restriction [75]. Other complications that become discernible later in the lives of these children include mental retardation, glaucoma, retinopathy, type 1 diabetes mellitus, thyroid dysfunction, and non-affective psychosis [76–81]. The anomalies that constitute CRS also contribute to miscarriages and infant mortality in the first 2 years of life [82].

With the implementation of an effective rubella vaccine program, CRS prevalence has become most significantly reduced in USA, continental Europe, Australia, and Japan; the reduced frequency in CRS is attributed to (1) a reduction in RV infection and (2) early termination of pregnancy when CRS is diagnosed. The other side of the issue however is that it continues to be a major health concern in countries where the rubella vaccine program is unavailable. A relatively recent estimate indicates that about 100,000 infants are born with CRS annually worldwide [83, 84]. With the advent of technological development and progressive emigration from countries without a routine rubella vaccine, there seems to be a global reemergence of the disease; however small the increase might be, CRS remains to be the most preventable congenital disease [85, 86].

4.2.1 Susceptible Periods of CRS

Rubella virus is known to cross the placenta [87–89]. Numerous studies have established the teratogenic effects of RV and the fact that most children with CRS have more than one defect [90–92]. The virus replicates in cell culture, disrupts cytoskeleton, impedes cell cycle, and causes mitochondrial changes and subsequently apoptosis of infected cells [93]. It infects and interacts with embryonic cells. It also slows down cell proliferation and induces apoptosis via complex mechanisms in infected organs. Both immunologic immaturity of the embryo and developmental stage are reported to influence the extent of teratogenic outcome following maternal primary rubella infection. The virus is reported to be present in the placenta during maternal viremia even before the rashes appear and the virus can be recovered from the placenta and from fetal tissues after therapeutic abortions, at birth in live fetuses and in about 10 % of cases at 1 year after birth. It causes necrosis of the trophoblast, progressive villus inflammation, and damage to endothelial cells of the chorionic vessels of the placenta [94, 95]. Microscopically vascular lesions in CRS patients are described as extensive local proliferation of tunica intima observed near the branching points [96, 97]. Rubella virus establishes persistent infection in the embryo, continuously replicating and instituting histopathological changes. Pathological modifications of fetal endothelial cells of the developing heart and great blood vessels are said to lead to cardiac septal defects, patent ductus arteriosus, and pulmonary stenosis. Intrauterine growth restriction (IUGR), congenital deafness, and neurodegenerative lesions are also described to result from RV-induced vascular insufficiency and fetal nutrient deprivation [98]. Continued presence of the virus and consequent fetal and placental vascular pathology might not only cause focal tissue damage and growth restriction [98, 99], but also complicates any conclusions drawn about critical periods of susceptibility to malformations in maternal rubella.

Reviewing the literature on clinical delineations of CRS, it becomes obvious that the early reports [71, 100–102] meticulously described the epidemiology, virus-induced embryo pathology, and phenotype of the infants; but their focus was not on sensitive periods. However, subsequent studies that looked into the temporal relationships of maternal rubella infection and congenital malformations of the offspring report different periods of susceptibilities [103]. For instance, De Santis et al. [104] summarized 35 reports of CRS cases into three trimesters with semester 1 being most susceptible, semester 2 moderately susceptible, and semester 3 only susceptible to IUGR. Based on their observations on a cohort 42 full-term infants with CRS with exact LMP date and time of onset of rubella rashes of the mothers during pregnancy, Ueda et al. [105] assumed that the critical period for rubella-associated IUGR was the interval from GD 16 to 100. Three most consistent anomalies observed in individuals with CRS are cardiac malformations, cataract, and deafness. A cutoff point of 16th–17th weeks appears to be repeatedly emphasized as the termination period of RA infection-associated fetal anomalies [103, 106, 107]. For example in Miller et al.s' study involving over 1000 women with confirmed rubella infection at different stages of development, pregnancy continued in 40 % and the infants were followed up clinically and serologically for 2 years [103]. Infants infected before the 11th week had mainly heart defects and deafness, whereas 35 % of those infected at 13–16 weeks had deafness alone. Infections after the 17th week caused no defects. Some of the cardiac anomalies and deafness do not manifest until later in life [108] and therefore these gestational time points do not necessarily define the commonly perceived critical periods of developmental susceptibility described in teratology and embryology texts. Also important to mention here is the secondary disruptive effects of the virus infection on the scala vestibuli, scala tympani, stria vascularis, hair cells of cochlea, modiolus, etc. on the subsequent development of the already formed inner ear components of the auditory organ system [109]. With respect to cataract, as in the case of congenital rubella-related deafness, it is difficult to obtain accurate estimate of the timing of embryonic insult by the virus. Since RV is blood-borne, it might not be able to enter the lens vesicle after the hyaloid blood vessels degenerate and lens capsule forms and therefore the risk of cataract should decline after week 5 of gestation [110]. However, this theory has been contested by Wolf [111] who asserts that the period of vulnerability to cataract extends to week 11. Ueda et al. [112] however observe that this susceptibility extends from 26 to 57 days (week 4–8) after fertilization. In vitro experiments on eye rudiments obtained from therapeutic abortions the rubella virus failed to cause cataracts when infected after the lens capsule enveloped the closed lens

vesicles (6–10-week embryos) [113]. Munro et al. [114] report CRS in infants following infection as late as 33 weeks. Development of the eye involves the processes of acquisition of precursors from diverse sources (optic vesicle from neurectoderm, lens placode from surface ectoderm, and the mesenchyme of neural crest origin), as well as morphogenetic processes such as cell proliferation, migration, differentiation, and apoptosis all which are tightly controlled by nuclear DNA-specific gene expression and epigenetic modifications [115, 116]. A similar complexity exists with respect to the otocyst and inner ear development [117, 118]. Therefore, it is obvious that the critical periods proposed for cataract and deafness in congenital rubella syndrome in the literature are based on often transient symptoms and/or patient history and they do not seem to take into account the chronic nature and severity of the infection and the molecular mechanisms of the embryology of the structures that are malformed. They also ignore the role of secondary effects of altered hemodynamics associated with concomitant vascular anomalies of CRS embryos. The temporal relationships between maternal rubella and fetal anomalies should be interpreted keeping in mind the progressive nature of the disease and the secondary effects that might mask any assumptions on the time of infection based on transient symptoms. What has also been overlooked further in the interpretation of the critical period in rubella teratogenesis is the fact that in addition to the developmental stage at which exposure to infectious agent occurs, there are several other factors that are contextually known to influence the teratogenic outcome (see Wilson [23]).

4.3 Retinoic Acid

4.3.1 Vitamin A Deficiency/Excess

In this review the term ***retinoid*** is used in place of vitamin A. The generic term *retinoid* encompasses both naturally occurring compounds with vitamin A activities and synthetic analogs of retinol with or without the biological activity [119]. Vitamin A derivatives include retinol, retinyl esters, all *trans*-retinoic acid, and its geometric isomer, 13-*cis* retinoic acid. Retinoids are known to control and maintain many biological processes such as development, normal growth of epithelial and skeletal structures, vision, and immune, reproductive, and integumentary systems [120]. Over 500 genes have been reported to be regulated by retinoic acid. Following ingestion, vitamin A is absorbed in the small intestine as a component of the chylomicrons and stored in liver parenchymal cells (80–90 %), and to a lesser extent in stellate cells and extraheptic adipocytes as retinyl esters or converted to retinal, the precursor of all-*trans*-retinoic acid. The 13-*cis* retinoic acid, also called retinoic acid, is *isotretinoin*. Several retinoids are used as therapeutic agents [121]. All currently available retinoids are teratogens [122]. Even before Greg [71] highlighted for the medical community the viral origin of birth defects, some landmark

experiments on the effect of maternal dietary restriction of vitamin A in pigs during pregnancy were shown to cause congenital anomalies in the offspring [123]. Pigs maintained on a diet vitamin A deficient supplemented with small amounts of cod liver oil for 160 days and then bred gave birth to offspring that had developed ocular, renal, gonadal, auricular, and palatal defects. Some of the surviving blind pigs when mated with normal pigs did not result in blindness in the offspring indicating that the anomalies were not heritable.

Warkany and his associates initiated a series of experiments in laboratory rats seeking to determine the fetal outcome of maternal dietary deficiencies during gestation [124–127]. These studies showed for the first time that the maternal vitamin A deficiency could induce a predictable pattern of fetal malformations which could be prevented by exogenous vitamin supplementation. Subsequently numerous reports in humans and animal models showed that excess vitamin A and its metabolites could produce a characteristic constellation of malformations that are dose and developmental stage specific [128–133]. Retinoic acid (RA-all *trans* retinoic acid), the metabolically most active metabolite of vitamin A, is now known to be a model teratogen and a thoroughly investigated morphogen, which regulates embryonic cell proliferation, differentiation, and apoptosis [134, 135].

5 Critical Period of Developmental Susceptibility to Retinoids

Numerous clinical case reports have reported a spectrum of craniofacial anomalies in infants following prenatal exposure to retinoids [136–143]. In the 16-month period following its release for the treatment of cystic acne approximately 120,000 women of childbearing age used isotretinoin. Since the deleterious fetal outcome became obvious, pregnancy prevention programs have been instituted for women of childbearing age group who intend to use isotretinoin. However, the drug continues to be used in pregnancy although significantly less frequently (0.32–0.95 per 1000) than before [144]. Retinoid use during pregnancy is reported to be associated with a high incidence of spontaneous abortions (about 20 %) as well as elective medical terminations to avoid birth of infants with several congenital malformations. Against a background malformation rate of 3 % in the population, isotretinoin-exposed pregnancies are reported to have a malformation rate, which is higher than that of thalidomide (35 % vs. 20 %) [145–148]. Isotretinoin-exposed pregnancies are further characterized by spontaneous abortions elective terminations [148]. The frequency of elective terminations is much higher than previously reported. There is robust evidence from clinical reports that retinoid exposure during gestation results in predictable pattern of

fetal malformations. Commonly observed anomalies include hydrocephalus, microcephalus, cerebellar hypoplasia, Fallot's tetralogy, transposition of great vessels, cardiac septal defects, aortic arch hypoplasia, thymic hypoplasia/aplasia, spina bifida, and limb defects. Whereas craniofacial malformations are suggestive of neural crest sensitivity to retinoids, several organs of non-neural origin are also affected in retinoid embryopathy. Children that appear physically normal with no gross malformations at birth are subsequently found to have mental retardation and impaired neuropsychological function [148, 149]. Children that appear physically normal with no gross malformations at birth are subsequently found to have mental retardation and impaired neuropsychological function [148]. Clinical observations have been corroborated by data from well-controlled laboratory animal experiments [131, 150–156]. Put in perspective these data would indicate that retinoic acid fulfills satisfactorily all the six Principles of Teratology that Wilson (1973) [23] enunciated. Clinical studies report that the critical period of maximum susceptibility to gross malformations in retinoid exposed pregnancies is weeks 2–5 postconception. However, one should remember that the pre-implantation period, the remaining organogenesis period, and the fetal period are not without risk [157, 158].

Retinoic acid (RA) alters embryogenesis in a developmental stage-dependent manner. It is important to note here that the amounts and periods of exposure to retinoids differ from pregnancy to pregnancy which should influence the fetal outcome differentially. Experiments in C3HRIxC57BL10/RI mice [159] indicate that all-*trans* retinoic acid (RA) administered at pre-implantation and early gastrulation stages (GD 4.5–5.5 postconception) results in supernumerary and ectopic limb induction, and limb and lower body duplications indicating that pre-implantation stages of mammalian development is susceptible to RA toxicity contrary to "all or none" refractory period concept [23, 28]. Experiments in TO mice show that doses higher than 75 mg/kg RA at pre-implantation stages result in complete failure of implantation. A lower dose (50 mg/kg) at GD 6.5–7.5 results in significant resorption and ectopic hindlimbs in the infraumbilical regions, often a single limb and absence of tail in about 4 % of the survivors (unpublished data). The susceptibility or lack of susceptibility thus appears to be dependent upon strain of the mice. In addition to the strain differences, there were methodological differences in experiments conducted in these two laboratories: whereas Rutledge et al. [159] suspended RA in DMSO and administered IP, our lab suspended RA (100 mg of RA) in a mixture of 0.5 ml ethanol plus 4.5 ml corn oil mixture and administered orally. DMSO possibly allows significantly slower rates of absorption of RA transperitoneally in contrast to the RA in ethanol-corn oil mixture via oral route, thus making RA available to interfere with differentiation of precursor cells of organ

primordia (caudal somites, hindlimb buds) which were not even present at the time of treatment. In fact RA as a teratogen is different from other teratogens in that it defies the developmental arrest theory by acting on the precursor cells that are destined to form organs whose primordia do not even exist at the point in time during development when it is administered. In the mouse, the forelimb buds appear as a thin ridge opposite somites 8–12 (late TS [Theiler Stage] 14, GD 9) when the embryo possesses about 15–20 somites and the hindlimb buds appear when the embryo has developed about 30–35 pairs of somites (TS 16, GD10) [160]. Various hindlimb malformation such as unipodia, sympodia, syrenomelia, and meromelia are induced in TO mouse embryos following a single dose of (100, 150, or 200 mg) RA administered on GD 8, when there are only up to seven pairs of somites and the limb buds are nonexistent [151]. These embryos also completely lacked tails and possessed severe malformations and/or agenesis of other caudal median structures at term [151].

Liao and Collins [161] observed hindlimb duplications in about 30 % of embryos of C57 mice following maternal treatment with RA on GD 5.5. They interpreted the anomalies in terms of the RA's ability to induce gene expression for ectopic primitive streaks, primitive nodes, and notochords as well as inhibition of endodermal and mesodermal cell migration. The visceral, skeletal, and neural structures that are affected in caudal dysplasia/regression syndrome in humans arise from ***caudal eminence*** (not to be confused with "tail bud" of chick embryos) first identified at Stage 9 between the cloacal membrane and neurenteric canal (Fig. 1 in O'Rahilly and Muller [162, 163]. Anomalies of the hindlimb, caudal neural tube, lower body wall, etc. observed in embryos treated with RA at GD 5.5 [151, 161] are a compelling experimental evidence that refutes the theory of developmental arrest [23] leading to malformation of an organ in which case the malformed organ would just resemble in appearance or mirror the stage when the teratogenic exposure occurred.

An additional fact that RA exemplifies is that certain organs have more than one phase of sensitivity to a teratogen. Experiments in Kochhar's laboratory on DBA/2J and ICR mice revealed two phases of maximum susceptibility to RA-induced cleft palate separated by a narrow window [164, 165]. However, this phenomenon appears to be species and strain dependent. The TO strain shows a very high incidence of cleft palate in response to RA administered on one of GD 8 through 13 lending support to a susceptibility continuum of cleft palate [166]. Although the end result (anomaly), namely cleft palate, remains the same, the pathogenetic mechanisms might be different for the same [167–169]. Retinoic acid (RA) and valproic acid (VPA) are both well-established teratogens, both in humans and laboratory animals. They both are known to induce anomalies of craniofacial structures of neural crest origin

and several organs of non-neural origin. A recent European study on maternal use of valproate monotherapy during first trimester of pregnancy reveals significant increases in risks for spina bifida, atrial septal defects, cleft palate, hypospadias, polydactyly, and craniosynostosis [170]. Short of structural malformations, VPA is also known to induce dose-dependent cognitive functional impairment in the offspring even after adjusting for common variables [171–173]. Maternal exposure to retinoic acid has a similar neurobehavioral effect on the offspring [174]. Although such behavioral anomalies could be attributed to the ability of the drug to alter gene expression, and inhibit neural and glial cell proliferation, synaptogenesis and apoptosis as evidenced from animal experiments, the exact mechanisms in humans remain unclear for now. Prenatal exposure to RA and antiepileptic agent valproic acid (VPA) also are known to cause intrauterine growth restriction (IUGR). Both VPA and RA are structurally weak acids with limited plasma half-life. However, when it comes to the question of critical periods for congenital malformations most studies suggest the first trimester as the most susceptible period. As a fat-soluble vitamin, about 80 % of consumed vitamin A is stored in the liver; thus even if the mother discontinues in-taking retinoids, the embryo continues to be exposed from the reserves. For obvious reasons treatment with VPA of epileptic pregnant women doesn't stop abruptly. Continued treatment with therapeutic agents might cause microscopic structural changes and suboptimal growth in mid and late gestational fetuses with lasting functional consequences. Therefore the first-semester critical period suggestion must be carefully interpreted.

Experimental studies indicate that there are malformation-specific critical periods with regard to RA and VPA teratogenesis. The results of Shenfelt's painstaking experiments with hamster [132, 133] illustrate clearly this malformation-specific critical periods for maternal RA treatment during gestation There are numerous examples in teratology that illustrate the fact that a certain malformation could be induced by different teratogens with significant differences in critical periods. For example, both RA and VPA induce vertebral and costal malformations and variations (supernumerary cervical and lumbar ribs and sternal anomalies) in mouse embryos. The vertebral anomalies include hemivertebrae, agenesis, fusion, as well as numerical reduction in presacral vertebrae [175]. The frequency and severity of these anomalies appear to be both dependent on developmental stage at teratogen administration and the strain of the mice experimented on. Thus in TO mouse fetuses a high frequency of cervical and lumbar ribs are induced by maternal treatment with VPA on GD 7 and GD 8, respectively. Fetuses with cervical ribs tend to have cervical neural arch anomalies, while fetuses with lumbar ribs often possess eight pairs of sternal ribs. In the RA experiments, the critical period for lumbar ribs is GD 8–12 while that for cervical ribs is GD 9–12. Peak incidence of

cervical ribs occurs following treatment on GD 10 and GD 11; lumbar rib incidence peaks are observed when treatment occurs on GD 8 and 11. The total number of presacral vertebrae is reduced from 26 to 25 and restricted to GD 7 (30 %) and GD 8 (18 %) of VPA treatment. The reduction is more severe (from 26 to as few as 23) in GD 8 VPA treatment. The critical period for RA-induced reduction in presacral vertebrae extends to GD 9. Thus VPA has a narrow window (GD 7 and 8) whereas RA has a longer window (GD 9–12) of susceptibility for the sternal defects. These axial skeletal malformations may be mediated by virtue of the ability of RA and VPA to alter gene expression [176, 177] or to induce histone deacetylation inhibition (HDACi) and epigenetic modification as suggested by Menegola et al. [178, 179].

6 Summary

In summary, it is apparent that birth defects are common, cause morbidity and mortality, and contribute significantly to economic burden. Preventative planning requires a thorough understanding of the pathogenetic mechanisms of birth defects. During the last 100 years the science of teratology has made remarkable progress in keeping pace in development in medical genetics, embryology, and developmental biology. Wilson postulated a set of principles of teratology which not only has stood the test of time but also deeply enhanced our understanding of teratogenic mechanisms. This paper looked at Wilson's principle of teratogenic susceptibility in the context of susceptible periods during development in relation to three model teratogens, namely ionizing radiation, congenital rubella, and retinoids. In view of the newer data in this area, it is becoming increasingly important to interpret the critical periods contextually, rather than in isolation. In other words, Wilson's six principles are interdependent. The whole is greater than the sum.

References

1. http://www.cdc.gov/ncbddd/birthdefects/infographic.html
2. Dolk H, Loane M, Garne E (2010) The prevalence of congenital anomalies in Europe. Adv Exp Med Biol 686:349–364
3. Ambe JP, Madziga AG, Akpede GO, Mava Y (2010) Pattern and outcome of congenital malformations in newborn babies in a Nigerian teaching hospital. West Afr J Med 29 (1):24–29
4. http://www.eurocat-network.eu
5. Yi Y, Lindemann M, Colligs A, Snowball C (2011) Economic burden of neural tube defects and impact of prevention with folic acid: a literature review. Eur J Pediatr 170 (11):1391–1400
6. Lo A, Polsek D, Siddu S (2014) Estimating the burden of neural tube defects in low- and middle-income countries. J Glob Health 4 (1):010402. doi:10.7189/jogh.04.010402
7. Colombo GL, Di Matteo S, Vinci M, Gatti C, Pascali MP, De Gennaro M, Macrellino E, Mosiello G, Redaelli T, Schioppa F, Dieci C (2013) A cost-of-illness study of spina bifida in Italy. Clinicoecon Outcomes Res 5: 309–316

8. Hertrampf E, Cortes F (2008) National food fortification program with folic acid in Chile. Food Nutr Bull 29(2 Suppl):S231–S237
9. Mosher WD, Jones J, Abma JC (2012) Intended and unintended births in the United States: 1982–2010. Natl Health Stat Report 55:1–28
10. Huynh L, McCoy M, Law A, Tran KN, Knuth S, Lefebvre P, Sullivan S, Duh MS (2013) Systematic literature review of the costs of pregnancy in the US. Pharmacoeconomics 31(11):1005–1030
11. Rabovskaja V, Parkinson B, Goodall S (2013) The cost-effectiveness of mandatory folic acid fortification in Australia. J Nutr 143 (1):59–66
12. Honein MA, Devine O, Grosse SD, Reefhuis J (2014) Prevention of orofacial clefts caused by smoking: implications of the Surgeon General's report. Birth Defects Res A Clin Mol Teratol. doi:10.1002/bdra.23274
13. Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller AB, Narwal R, Adler A, Vera Garcia C, Rohde S, Say L, Lawn JE (2012) National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet 379(9832):2162–2172
14. Bortolus R, Blom F, Filippini F, van Poppel MN, Leoncini E, de Smit DJ, Benetollo PP, Cornel MC, de Walle HE, Mastroiacovo P (2014) Italian and Dutch folic acid trial study groups. Prevention of congenital malformations and other adverse pregnancy outcomes with 4.0 mg of folic acid: community-based randomized clinical trial in Italy and the Netherlands. BMC Pregnancy Childbirth 14:166. doi:10.1186/1471-2393-14-166
15. Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller AB, Kinney M, Lawn J, Born Too Soon Preterm Birth Action Group (2013) Born too soon. The global epidemiology of 15 million preterm births. Reprod Health 10(Suppl 1):S2. doi:10.1186/1742-4755-10-S1-S2
16. Melve KK, Skjaerven R (2002) Families with birth defects: is birth weight of nonmalformed siblings affected? Am J Epidemiol 155(10):932–940
17. Bulkowski R (2010) Stillbirth and fetal growth restriction. Clin Obstet Gynecol 53 (3):673–680
18. de Onis M, Blossner M, Villar J (1998) Levels and patterns of intrauterine growth retardation in developing countries. Eur J Clin Nutr 52(Suppl):S5–S15
19. Gluckman PD, Hanson MA, Cooper C, Thornburg KL (2008) Effect of in utero and early-life conditions on adult health and disease. N Engl J Med 359(1):61–73
20. Gluckman PD, Hanson MA (2008) Developmental and epigenetic pathways to obesity: an evolutionary-developmental perspective. Int J Obes (Lond) 32(Suppl 7):S62–S71
21. Fall CH (2013) Fetal programming and the risk of noncommunicable disease. Indian J Pediatr 80(Suppl 1):S13–S20
22. Fall CH (2013) Fetal malnutrition and long-term outcomes. Nestle Nutr Inst Workshop Ser 74:11–25, Epub 2013 Jul 18
23. Wilson JG (1973) Environment and birth defects. Academic, New York
24. Jelínek R (2005) The contribution of new findings and ideas to the old principles of teratology. Reprod Toxicol 20(3):295–300
25. Charon P (2004) Etienne Geoffroy Saint-Hilaire (1772–1844) and anencephaly: contribution of one naturalist to medical knowledge. Hist Sci Med 38(3):365–383
26. https://archive.org/details/Geoffroy-Saint-Hilaire1832xs96J-c.
27. Dareste C (1877) Recherches sur la production artificielle des monstruosités: ou, Essais de tératogénie expérimentale [Research on the artificial production of monstrosities, or experimental teratogenicity testing]. C. Reinwald, Paris, http://www.biodiversitylibrary.org/item/99425#page/48/mode/1up
28. Stockard CR (1921) Developmental rate and structural expression: an experimental study of twins, double monsters and single deformities, and the interaction among embryonic organs during their origin and development. Am J Anat 28:115–227
29. Ujházy E, Mach M, Navarová J, Brucknerová I, Dubovický M (2012) Teratology—past, present and future. Interdiscip Toxicol 5 (4):163–168
30. http://www.slideshare.net/khushnood18/biographical-encyclopedia-of-scientists-3rd-ed-malestrom
31. Lenz W (1985) Thalidomide embryopathy in Germany, 1959–1961. Prog Clin Biol Res 163C:77–83
32. Lenz W (1988) A short history of thalidomide embryopathy. Teratology 38(3):203–215
33. Williams PM, Fletcher S (2010) Health effects of prenatal radiation exposure. Am Fam Physician 82(5):488–493

34. Jelínek R (1988) The principles of teratogenesis revisited. Congenit Anom 28:S145–S155
35. Friedman JM (2010) The principles of teratology: are they still true? Birth Defects Res A Clin Mol Teratol 88(10):766–768
36. Holliday R (1998) The possibility of epigenetic transmission of defects induced by teratogens. Mutat Res 422(2):203–205
37. Ema M, Ise R, Kato H, Oneda S, Hirose A, Hirata-Koizumi M, Singh AV, Knudsen TB, Ihara T (2010) Fetal malformations and early embryonic gene expression response in cynomolgus monkeys maternally exposed to thalidomide. Reprod Toxicol 29(1):49–56
38. Wilson JG (1959) Experimental studies on congenital malformations. J Chronic Dis 10:111–130
39. Stockard CR (1928) Present status of the problem of the so-called rejuvenation. Bull N Y Acad Med 4(12):1241–1249
40. Goldstein L, Murphy DP (1929) Microcephalic idiocy following radium therapy for uterine cancer during pregnancy. Am J Obstet Gynecol 18:189–195, 289–3
41. Goldstein L, Murphy DP (1929) Etiology of ill health in children born after maternal pelvic irradiation. Part 11. Defective children born after postconceptional maternal irradiation. Am J Roentgenol 22:322–331
42. Doll R (1995) Hazards of ionizing radiation: 100 years of observation on man. Br J Cancer 72(6):1339–1349
43. Plummer G (1952) Anomalies occurring in children exposed in utero to atomic bomb in Hiroshima. Pediatrics 10(6):687–693
44. Dekaban AS (1968) Abnormalities in children exposed to X radiation during various stages of gestation: tentative timetable of radiation injury to the human fetus I. J Nucl Med 9 (9):471–477
45. Otake M, Schull WJ, Yoshimaru H (1991) A review of forty-five years study of Hiroshima and Nagasaki atomic bomb survivors. Brain damage among the prenatally exposed. J Radiat Res 32(Suppl):249–264
46. Hall EJ (1991) Scientific view of low-level radiation risks. Radiographics 11(3):509
47. De Santis M, Di Gianantonio E, Straface G, Cavaliere AF, Caruso A, Schiavon F, Berletti R, Clementi M (2005) Ionizing radiations in pregnancy and teratogenesis: a review of literature. Reprod Toxicol 20(3):323–329
48. Schull WJ, Otake M (1999) Cognitive function and prenatal exposure to ionizing radiation. Teratology 59:222–226
49. Dobbing J (1974) The later growth of the brain and its vulnerability. Pediatrics 53 (1):2–6
50. Janeczko K, Setkowicz Z, Fraczek M, Kochowska J (1999) Effects of prenatal gamma-irradiation on postnatal astrogliogenesis in the hippocampal formation of rat. Brain Res 816(2):628–632
51. Bruni JE, Persaud TV, Huang W, Froese G (1993) Postnatal development of the rat CNS following in utero exposure to a low dose of ionizing radiation. Exp Toxicol Pathol 45 (4):223–231
52. D'Amato CJ, Hicks SP (1965) Effects of low levels of ionizing radiation on the developing cerebral cortex of the rat. Neurology 15 (12):1104–1116
53. Hicks SP, D'Amato CJ (1980) Effects of radiation on development, especially of the nervous system. Am J Forensic Med Pathol 1 (4):309–317
54. Dagg CP (1966) Teratogenesis. In: Green EL (ed) Biology of the laboratory mouse, 2nd edn. McGraw-Hill, New York, pp 309–328
55. Dobbing J, Sands J (1973) Quantitative growth and development of human brain. Arch Dis Child 48:757–767
56. Clancy B, Kersh B, Hyde J, Darlington RB, Anand KJ, Finlay BL (2007) Web-based method for translating neurodevelopment from laboratory species to humans. Neuroinformatics 5(1):79–94
57. Clancy B, Finlay BL, Darlington RB, Anand KJ (2007) Extrapolating brain development from experimental species to humans. Neurotoxicology 28(5):931–937
58. Chugani HT (1998) A critical period of brain development: studies of cerebral glucose utilization with PET. Prev Med 27(2):184–188
59. Rice D, Barone S Jr (2000) Critical periods of vulnerability for the developing nervous system: evidence from humans and animal models. Environ Health Perspect 108(Suppl 3):511–533
60. Ratnapalan S, Bona N, Chandra K, Koren G (2004) Physicians' perceptions of teratogenic risk associated with radiography and CT during early pregnancy. AJR Am J Roentgenol 182(5):1107–1109
61. Sodickson A (2012) Strategies for reducing radiation exposure in multi-detector row CT. Radiol Clin North Am 50(1):1–14
62. Smith-Bindman R, Miglioretti DL, Johnson E, Lee C, Feigelson HS, Flynn M, Greenlee RT, Kruger RL, Hornbrook MC, Roblin D, Solberg LI, Vanneman N, Weinmann S, Williams AE (2012) Use of diagnostic imaging

studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996–2010. JAMA 307 (22):2400–2409

63. Hudson MM (2010) Reproductive outcomes for survivors of childhood cancer. Obstet Gynecol 116(5):1171–1183
64. Signorella LB, Cohen SS, Bosetti C, Stovall M, Kasper CE, Weathers RE, Whitton JA, Green DM, Donaldson SS, Mertens AC, Robison LL, Boice JD Jr (2006) Female survivors of childhood cancer: preterm birth and low birth weight among their children. J Natl Cancer Inst 98(20):1453–1461
65. Brent RL, Beckman DA (1990) Environmental teratogens. Bull N Y Acad Med 66 (2):123–163
66. Tank KL (1987) Intra-uterine infections. Ann Acad Med Singapore 16(4):707–712
67. Collier SA, Rasmussen SA, Feldkamp ML, Honein MA (2009) Prevalence of self-reported infection during pregnancy among control mothers in the National Birth Defects Prevention Study. National Birth Defects Prevention Study. Birth Defects Res A Clin Mol Teratol 85(3):193–201
68. Menser MA, Forrest JM (1974) Rubella—high incidence of defects in children considered normal at birth. Med J Aust 1 (5):123–126
69. Lee JY, Bowden DS (2000) Rubella virus replication and links to teratogenicity. Clin Microbiol Rev 13(4):571–587
70. Wesselhoeft C (1947) Rubella (German measles). N Engl J Med 236:943–950
71. Forbes JA (1969) Rubella: historical aspects. Am J Dis Child 118(1):5–11
72. Gregg NM (1991) Congenital cataract following German measles in the mother. 1941. Aust N Z J Ophthalmol 19(4):267–276
73. Cooper LZ, Alford CA Jr (2006) Rubella, Chapter 28. In: Remington JS, Klein JO, Baker CJ, Wilson CB (eds) Infectious diseases of the fetus and newborn infant, 6th edn. Elsevier, Philadelphia, pp 893–926
74. http://www.cdc.gov/mmwr/preview/mmwr html/rr6204a1.htm
75. Ghidini A, Lynch L (1993) Prenatal diagnosis and significance of fetal infections. West J Med 159(3):366–373
76. Forester JM, Menser MA, Burgess JA (1971) High frequency of diabetes mellitus in young adults with congenital rubella. Lancet 2 (7720):332–334
77. Johnson GM, Tudor RB (1970) Diabetes mellitus and congenital rubella infection. Am J Dis Child 120(5):453–455
78. Givens KT, Lee DA, Jones T, Ilstrup DM (1993) Congenital rubella syndrome: ophthalmic manifestations and associated systemic disorders. Br J Ophthalmol 77 (6):358–363
79. Arnold JJ, McIntosh ED, Martin FJ, Menser MA (1994) A fifty-year follow-up of ocular defects in congenital rubella: late ocular manifestations. Aust N Z J Ophthalmol 22(1):1–6
80. O'Neill JF (1998) The ocular manifestations of congenital infection: a study of the early effect and long-term outcome of maternally transmitted rubella and toxoplasmosis. Trans Am Ophthalmol Soc 96:813–879
81. Vijayalakshmi P, Kakkar G, Samprathi A, Banushree R (2002) Ocular manifestations of congenital rubella syndrome in a developing country. Indian J Ophthalmol 50:307–311
82. Hansha JB, Dudgeon JB, Marshall WC (1985) Viral disease of the fetus and newborn, 2nd edn. Saunders, Philadelphia
83. Robertson SE, Featherstone DA, Gacic-Dobo M, Hersh BS (2003) Rubella and congenital rubella syndrome: global update. Pan Am J Public Health 14(5):306–315
84. McLean H, Red S, Abernathy E, Icenogle J, Wallace G (2012) Congenital rubella syndrome: Chapter 15-1. In: VPD surveillance manual, 5th edn. CDC, Atlanta, GA, http://www.cdc.gov/vaccines/pubs/surv-manual/chpt15-crs.pdf
85. Smits G, Mollema L, Hahné S, de Melker H, Tcherniaeva I, van der Klis F, Berbers G (2014) Seroprevalence of rubella antibodies in The Netherlands after 32 years of high vaccination coverage. Vaccine 32 (16):1890–1895
86. Hutton J, Rowan P, Greisinger A, Mouzoon M (2014) Rubella monitoring in pregnancy as a means for evaluating a possible reemergence of rubella. Am J Obstet Gynecol 211(5):534.e1–534.e4
87. Fleet WF Jr, Benz EW Jr, Karzon DT, Lefkowitz LB, Herrmann KL (1974) Fetal consequences of maternal rubella immunization. JAMA 227(6):621–627
88. Wyll SA, Herrmann KL (1973) Inadvertent rubella vaccination of pregnant women. Fetal risk in 215 cases. JAMA 225(12):1472–1476
89. Bolognese RJ, Corson SL, Fuccillo DA, Sever JL, Traub R (1973) Evaluation of possible transplacental infection with rubella vaccination during pregnancy. Am J Obstet Gynecol 117(7):939–941
90. Dudgeon JA (1970) Teratogenic effect of rubella virus. Proc R Soc Med 63 (12):1254–1257

91. Cooper LZ (1985) The history and medical consequences of rubella. Rev Infect Dis 7 (Suppl 1):S2–S10
92. Cutts FT, Robertson SE, Diaz-Ortega JL, Samuel R (1997) Control of rubella and congenital rubella syndrome (CRS) in developing countries, Part 1: Burden of disease from CRS. Bull World Health Organ 75(1):55–68
93. Atreya CD, Mohan KV, Kulkarni S (2004) Rubella virus and birth defects: molecular insights into the viral teratogenesis at the cellular level. Birth Defects Res A Clin Mol Teratol 70(7):431–437
94. Töndury G, Smith DW (1966) Fetal rubella pathology. J Pediatr 68(6):867–879
95. Driscoll SG (1969) Histopathology of gestational rubella. Am J Dis Child 118(1):49–53
96. Varghese PJ, Izukawa T, Rowe RD (1969) Supravalvular aortic stenosis as part of rubella syndrome, with discussion of pathogenesis. Br Heart J 31(1):59–62
97. Vince DJ (1970) The role of rubella in the etiology of supravalvular aortic stenosis. Can Med Assoc J 103(11):1157–1160
98. Perelygina L, Zheng Q, Metcalfe M, Icenogle J (2013) Persistent infection of human fetal endothelial cells with rubella virus. PLoS One 8(8):e73014, Epub 2013 Aug 5
99. Frey TK (1994) Molecular biology of rubella virus. Adv Virus Res 44:69–160
100. Horstmann DM (1965) Rubella and the rubella syndrome: new epidemiologic and virologic observations. Calif Med 102:397–403
101. Horstmann DM, Banatvala JE, Riordan JT, Payne MC, Whittemore R, Opton EM, duVe Florey C (1965) Maternal rubella and the rubella syndrome in infants. Epidemiologic, clinical, and virologic observations. Am J Dis Child 110(4):408–415
102. Banatvala JE, Horstmann DM, Payne MC, Gluck L (1965) Rubella syndrome and thrombocytopenic purpura in newborn infants. Clinical and virological observations. N Engl J Med 273(9):474–478
103. Miller E, Cradock-Watson JE, Pollock TM (1982) The consequences of confirmed maternal rubella at successive stages of pregnancy. Lancet 2(8302):781–784
104. De Santis M, Cavaliere AF, Straface G, Caruso A (2006) Rubella infection in pregnancy. Reprod Toxicol 21(4):390–398
105. Ueda K, Hisanaga S, Nishida Y, Shepard TH (1981) Low-birth-weight and congenital rubella syndrome: effect of gestational age at time of maternal rubella infection. Clin Pediatr (Phila) 20(11):730–733
106. Ueda K, Nishida Y, Oshima K, Shepard TH (1979) Congenital rubella syndrome: correlation of gestational age at time of maternal rubella with type of defect. J Pediatr 94 (5):763–765
107. Tokugawa K, Ueda K (1986) Rubella epidemicity and endemicity in Japan: analysis of birth years of the 365 cases of deaf children with history of maternal rubella. Int J Epidemiol 15(4):585–587
108. Best JM (2007) Rubella. Semin Fetal Neonatal Med 12(3):182–192
109. Friedman I, Wright MI (1966) Histopathological changes in the foetal and infantile inner ear caused by maternal rubella. Br Med J 2 (5504):20–23
110. Boniuk M, Zimmerman LE (1967) Ocular pathology in the rubella syndrome. Arch Ophthalmol 77:455–473
111. Wolf SM (1985) Rubella syndrome. In: Darrell RW (ed) Viral diseases of the eye. Lea & Febiger, Philadelphia, pp 199–207
112. Ueda K, Tokugawa K, Kusuhara K (1992) Perinatal viral infections. Early Hum Dev 29 (1–3):131–135
113. Karkinen-Jääskeläinen M, Saxén L, Vaheri A, Leinikki P (1975) Rubella cataract in vitro: sensitive period of the developing human lens. J Exp Med 141(6):1238–1248
114. Munro ND, Sheppard S, Smithells RW, Holzel H, Jones G (1987) Temporal relations between maternal rubella and congenital defects. Lancet 2(8552):201–204
115. Cvkl A, Duncan MK (2007) Genetic and epigenetic mechanisms of gene regulation during lens development. Prog Retin Eye Res 26 (6):555–597
116. Yan B, Yao J, Tao ZF, Jiang Q (2014) Epigenetics and ocular diseases: from basic biology to clinical study. Cell Physiol 229(7):825–833
117. Weston MD, Pierce ML, Rocha-Sanchez S, Beisel KW, Soukup GA (2006) MicroRNA gene expression in the mouse inner ear. Brain Res 1111(1):95–104
118. Quint E, Steel KP (2003) Use of mouse genetics for studying inner ear development. Curr Top Dev Biol 57:45–83
119. Goldman DS (1984) Vitamin A, and retinoids in health and disease. N Engl J Med 310:1023–1031
120. D'Ambrosio DN, Clugston RD, Blaner WS (2011) Vitamin A metabolism: an update. Nutrients 3(1):63–103
121. Hinds TS, West WL, Knight EM (1997) Carotenoids and retinoids: a review of research,

clinical, and public health applications. J Clin Pharma 37:551–558

122. Digiovanna JJ, Mauro T, Milstone LM, Schmuth M, Toro JR (2013) Systemic retinoids in the management of ichthyoses and related skin types. Dermatol Ther 26 (1):26–38
123. Hale F (1933) Pigs born without eyeballs. J Hered 24:105–106
124. Warkany J, Nelson RC (1940) Appearance of skeletal abnormalities in the offspring of rats reared on a deficient diet. Science 92 (2391):383–384
125. Warkany J, Schraffenberger E (1946) Congenital malformations induced in rats by maternal vitamin A deficiency defects of the eye. Arch Ophthal 35:150–169
126. Wilson JG, Warkany J (1947) Anomalies of the genito-urinary tract induced by maternal vitamin A deficiency in fetal rats. Anat Rec 97 (3):376
127. Warkany J (1948) Experimental studies on nutrition in pregnancy. Obstet Gynecol Surv 3(5):693–703
128. Wilson JG, Roth CB, Warkany J (1953) An analysis of the syndrome of malformations induced by maternal vitamin A deficiency. Effect of restoration of vitamin A at various times during gestation. Am J Anat 92:189–217
129. Cohlan SQ (1953) Excessive intake of vitamin A as a cause of congenital anomalies in the rat. Science 117(3046):535–536
130. Cohlan SQ (1954) Congenital anomalies in the rat produced by excessive intake of vitamin A during pregnancy. Pediatrics 13 (6):556–567
131. Kochhar DM (1967) Teratogenic activity of retinoic acid. Acta Pathol Microbiol Scand 70 (3):398–404
132. Shenefelt RE (1972) Morphogenesis of malformations in hamsters caused by retinoic acid: relation to dose and stage at treatment. Teratology 5(1):103–118
133. Shenefelt RE (1972) Gross congenital malformations. Animal model: treatment of various species with a large dose of vitamin A at known stages in pregnancy. Am J Pathol 66 (3):589–592
134. Maden M (2001) Vitamin A, and the developing embryo. Postgrad Med J 77 (910):489–491
135. Niederreither K, Subbarayan V, Dollé P, Chambon P (1999) Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. Nat Genet 21 (4):444–448
136. Adams J, Lammer EJ (1993) Neurobehavioral teratology of isotretinoin. Reprod Toxicol 7(2):175–177
137. Rizzo R, Lammer EJ, Parano E, Pavone L, Argyle JC (1991) Limb reduction defects in humans associated with prenatal isotretinoin exposure. Teratology 44(6):599–604
138. Lynberg MC, Khoury MJ, Lammer EJ, Waller KO, Cordero JF, Erickson JD (1990) Sensitivity, specificity, and positive predictive value of multiple malformations in isotretinoin embryopathy surveillance. Teratology 42(5):513–519
139. Werler MM, Lammer EJ, Rosenberg L, Mitchell AA (1990) Maternal vitamin A supplementation in relation to selected birth defects. Teratology 42(5):497–503
140. Lammer EJ, Schunior A, Hayes AM, Holmes LB (1988) Isotretinoin dose and teratogenicity. Lancet 2(8609):503–504
141. Lammer EJ, Hayes AM (1988) Isotretinoin phenocopy. Am J Med Genet 29(3):675–678
142. Lammer EJ, Chen DT, Hoar RM, Agnish ND, Benke PJ, Braun JT, Curry CJ, Fernhoff PM, Grix AW Jr, Lott IT et al (1985) Retinoic acid embryopathy. N Engl J Med 313 (14):837–841
143. Lammer EJ, Flannery DB, Barr M (1985) Does isotretinoin cause limb reduction defects? Lancet 2(8450):328
144. Rouzès A, Jonville-Béra AP (2014) Exposure to isotretinoin during pregnancy in France: 25 years of follow-up. Therapie 69(1):53–63
145. Stern RS, Rosa F, Baum C (1984) Isotretinoin and pregnancy. Am Acad Dermatol 10(5 Pt 1):851–854
146. Dai WS, LaBraico JM, Stern RS (1992) Epidemiology of isotretinoin exposure during pregnancy. J Am Acad Dermatol 26 (4):599–606
147. Newman CG (1985) Teratogen update: clinical aspects of thalidomide embryopathy—a continuing preoccupation. Teratology 32 (1):133–144
148. Sladden MJ, Harman KE (2007) What is the chance of a normal pregnancy in a woman whose fetus has been exposed to isotretinoin? Arch Dermatol 143(9):1187–1188
149. Bérard A, Azoulay L, Koren G, Blais L, Perreault S, Oraichi D (2007) Isotretinoin, pregnancies, abortions and birth defects: a population-based perspective. Br J Clin Pharmacol 63(2):196–205
150. Webster WS, Johnston MC, Lammer EJ, Sulik KK (1986) Isotretinoin embryopathy and the cranial neural crest: an in vivo and in vitro study. J Craniofac Genet Dev Biol 6 (3):211–222

151. Padmanabhan R (1998) Retinoic acid-induced caudal regression syndrome in the mouse fetus. Reprod Toxicol 12:139–151
152. Yasuda Y, Konishi H, Kihara T, Tanimura T (1990) Discontinuity of primary and secondary neural tube in spina bifida induced by retinoic acid in mice. Teratology 41:257–274
153. Quemelo PR, Lourenço CM, Peres LC (2007) Teratogenic effect of retinoic acid in swiss mice. Acta Cir Bras 22(6):451–456
154. Troncoso Sch M, Rojas HC, Bravo CE (2008) Isotretinoin embryopathy. Report of one case. Rev Med Chil 136(6):763–766
155. Yu J, Gonzalez S, Diez-Pardo JA, Tovar JA (2002) Effects of vitamin A on malformations of neural-crest-controlled organs induced by nitrofen in rats. Pediatr Surg Int 18 (7):600–605
156. Hart RC, McCue PA, Ragland WL, Winn KJ, Unger ER (1990) Avian model for 13-cis-retinoic acid embryopathy: demonstration of neural crest related defect. Teratology 41 (4):463–472
157. Huang FJ, Lin YL (2001) Effects of retinoic acid on pre-implantation embryo development in mice. Chang Gung Med J 24 (11):681–687
158. Huang FJ, Hsuuw YD, Lan KC, Kang HY, Chang SY, Hsu YC, Huang KE (2006) Adverse effects of retinoic acid on embryo development and the selective expression of retinoic acid receptors in mouse blastocysts. Hum Reprod 21(1):202–209
159. Rutledge JC, Shourbaji AG, Hughes LA, Polifka JE, Cruz YP, Bishop JB, Generoso WM (1994) Limb and lower-body duplications induced by retinoic acid in mice. Proc Natl Acad Sci U S A 91(12):5436–5440
160. Kaufman MH, Bard JBL (1999) The anatomical basis of mouse development. Academic, San Diego, CA, p 33
161. Liao X, Collins MD (2008) All-trans retinoic acid-induced ectopic limb and caudal structures: murine strain sensitivities and pathogenesis. Dev Dyn 237(6):1553–1564
162. O'Rahilly R, Müller F (1989) Interpretation of some median anomalies as illustrated by cyclopia and symmelia. Teratology 40 (5):409–421
163. Padmanabhan R, Naruse I, Shiota K (1999) Caudal dysgenesis in staged human embryos: Carnegie stages 16–23. Am J Med Genet 87 (2):115–127
164. Kochhar DM (1973) Limb development in mouse embryos. I. Analysis of teratogenic effects of retinoic acid. Teratology 7 (3):289–295
165. Kochhar DM, Penner JD, Tellone CI (1984) Comparative teratogenic activities of two retinoids: effects on palate and limb development. Teratog Carcinog Mutagen 4 (4):377–387
166. Padmanabhan R, Ahmed I (1997) Retinoic acid-induced asymmetric craniofacial growth and cleft palate in the TO mouse fetus. Reprod Toxicol 11(6):843–860
167. Abbott BD, Hill LG, Birnbaum LS (1990) Processes involved in retinoic acid production of small embryonic palatal shelves and limb defects. Teratology 41(3):299–310
168. Abbott BD, Birnbaum LS (1990) Retinoic acid-induced alterations in the expression of growth factors in embryonic mouse palatal shelves. Teratology 42(6):597–610
169. Abbott BD, Harris MW, Birnbaum LS (1989) Etiology of retinoic acid-induced cleft palate varies with the embryonic stage. Teratology 40(6):533–553
170. Jentink J, Loane MA, Dolk H, Barisic I, Garne E, Morris JK, de Jong-van den Berg LT, EUROCAT Antiepileptic Study Working Group (2010) Valproic acid monotherapy in pregnancy and major congenital malformations. N Engl J Med 362(23):2185–2193
171. Meador KJ, Baker GA, Browning N, Cohen MJ, Bromley RL, Clayton-Smith J, Kalayjian LA, Kanner A, Liporace JD, Pennell PB, Privitera M, Loring DW, NEAD Study Group (2013) Fetal antiepileptic drug exposure and cognitive outcomes at age 6 years (NEAD study): a prospective observational study. Lancet Neurol 12(3):244–252
172. Bromley R, Weston J, Adab N, Greenhalgh J, Sanniti A, McKay AJ, Tudur Smith C, Marson AG (2014) Treatment for epilepsy in pregnancy: neurodevelopmental outcomes in the child. Cochrane Database Syst Rev 10, CD010236. doi:10.1002/14651858.CD010236.pub2
173. Ornoy A (2006) Neuroteratogens in man: an overview with special emphasis on the teratogenicity of antiepileptic drugs in pregnancy. Reprod Toxicol 22(2):214–226
174. Adams J (1993) Structure-activity and dose-response relationships in the neural and behavioral teratogenesis of retinoids. Neurotoxicol Teratol 15(3):193–202
175. Rengasamy P, Padmanabhan RR (2004) Experimental studies on cervical and lumbar

ribs in mouse embryos. Congenit Anom (Kyoto) 44(3):156–171

176. Kessel M (1992) Respecification of vertebral identities by retinoic acid. Development 115 (2):487–501
177. Kawanishi CY, Hartig P, Bobseine KL, Schmid J, Cardon M, Massenburg G, Chernoff N (2003) Axial skeletal and Hox expression domain alterations induced by retinoic acid, valproic acid, and bromoxynil during murine development. J Biochem Mol Toxicol 17(6):346–356
178. Menegola E, Di Renzo F, Broccia ML, Prudenziati M, Minucci S, Massa V, Giavini E (2005) Inhibition of histone deacetylase activity on specific embryonic tissues as a new mechanism for teratogenicity. Birth Defects Res B Dev Reprod Toxicol 74 (5):392–398
179. Giavini E, Menegola E (2014) Teratogenic activity of HDAC inhibitors. Curr Pharm Des 20(34):5438–5442

Methods in Pharmacology and Toxicology (2017) 43–115
DOI 10.1007/7653_2015_53
© Springer Science+Business Media New York 2015
Published online: 10 April 2016

Nonclinical Safety Assessment of Developmental and Reproductive Toxicology: Considerations for Conducting Fertility, Embryo–Fetal Development, and Prenatal and Postnatal Developmental Toxicology Studies

Kevin H. Denny and Ali S. Faqi

Abstract

This chapter reviews the current regulatory considerations for reproductive and developmental toxicology testing specifically relating to pharmaceuticals and biopharmaceuticals. A brief history of the regulatory perspective of drugs which provides background to the current study designs presented in the chapter is discussed. The initial FDA guidelines and the subsequent International Conference on Harmonisation (ICH) S5(R2) currently used for the assessment of potential developmental and reproductive toxicity for safety evaluations of medicinal products are in some measure a testament of the endeavor of government authorities to prevent future tragedies, such as that of Thalidomide. The chapter incorporates reviews of the basic requirements for Conducting Fertility, Embryo–Fetal Development, and Prenatal and Postnatal Developmental Toxicology Studies as core components of pharmaceutical drug development programs. With the increase in biopharmaceutical development, the chapter also includes consideration for evaluating Reproductive and Developmental Toxicity in nontraditional animal models. Alternative methods for in vitro evaluations of reproductive and developmental toxicology are also reviewed in this chapter.

Keywords: Fertility and early embryonic development, Embryo–fetal development, Prenatal and postnatal developmental toxicity, Dose selection, Maternal observations, Postweaning assessments, Nonhuman primate, Juvenile toxicity studies, Embryonic stem cell test, Whole embryo culture, Zebrafish embryotoxicity test

1 Overview and Historical Background of Regulations and Guidelines for Reproductive and Developmental Toxicology Assessments

1.1 Background and Historical Perspective

The current regulatory environment relating to safety evaluations in the nonclinical development of potential new pharmaceuticals is the result of a tragedy experienced by thousands of families worldwide induced by Thalidomide. Thalidomide first developed by Chemie Grünenthal in West Germany was launched as Contergan (Thalidomide) in 1957 and by 1960 Thalidomide was being marketed by at least 14 firms in many countries under 37 different trade names and was even sold in many countries without a prescription [1].

The manufacturer and distributors touted Thalidomide as safe and effective for numerous conditions including nausea and morning sickness during pregnancy. Widespread use of Thalidomide resulted in fetal deaths and deformities in children born after their mothers used it during pregnancy and by November 1961, thalidomide was taken off the market due to massive pressure from the press and public. It was later found to be teratogenic in fetal development, most visibly as a cause of amelia or phocomelia as the drug is a angiogenesis inhibitor—interfering with blood vessel development, especially if taken during the first 25–50 days of pregnancy [2, 3]. In the late 1950s and early 1960s, more than 10,000 children in 46 countries were born with deformities such as phocomelia as a consequence of thalidomide use [2]. It is not known exactly how many worldwide victims of the drug there have been, although estimates range from 10,000 to 20,000 [1].

Despite of the reports of major birth defects being attributed to Thalidomide in Europe, Japan and Australia, the William S. Merrill Company attempted to market Thalidomide as a sedative in the USA. In September 1960 the FDA received a New Drug Application from Merrell that requested approval for thalidomide. The NDA provided by Merrell contained data from previous animal and human studies indicating at the therapeutic human dosage it was both safe and effective [4]. Since thalidomide was already widely used in many other countries (presumably), approval of the NDA was thought to be routine and was assigned to the agency's newest medical officer, Dr. Frances Oldham Kelsey. Kelsey later explained that "They gave it [the NDA for thalidomide] to me because they thought it would be an easy one to start on. As it turned out, it wasn't all that easy." [5] Under the laws in place at that time, Merrell would have been able to market thalidomide if it had not heard back from the FDA within 60 days of submitting its application. Therefore, if Dr. Kelsey had done nothing, the drug would have been marketed. Indeed, Merrell was poised to begin distribution in the USA having brought in at least 5 t of the drug to its warehouses [6]. Fortunately the company was prevented by FDA medical officer Frances Kelsey, Ph.D., M.D., who refused to approve the drug application because of insufficient safety data, with the submitted evidence being more anecdotal than clinical.

Five months after the NDA was submitted, in February 1961, a letter published in the British Medical Journal reported the possible occurrence of peripheral neuritis (deterioration of the nerves in the hands and feet) in patients who had used thalidomide over a long term. While the NDA was never approved Thalidomide the FDA discovered that many of the doctors had received the drug from Merrell and that more than 2,500,000 tablets had been distributed to over 20,000 patients of whom 624 were reported to have been pregnant. Eventually well-documented cases were found where women who had received thalidomide in the USA during pregnancy delivered seriously deformed babies [5].

During the time that the Merrell Company was seeking approval for the drug, they stated that the drug had been so widely used in Europe that any adverse effects on a fetus would have been reported. It turned out, however, that a group at Bonn University in Germany had noticed an increased incidence of babies born in their hospital with badly deformed extremities, as early as 1959. They learned that a similar increase in such deformities was reported in England and in Sweden. The Bonn investigators suspected thalidomide might be associated with the deformities but gave up the idea after finding out that similar increases in birth defects had not been noted in the USA or Canada. The Bonn researchers had assumed that thalidomide was available in the USA based on statements in the promotional literature of the German firm, and so did not realize that thalidomide was indeed the culprit [7].

1.2 Early Guidelines

The tragedy of thalidomide led to changes that strengthened both the regulatory and scientific environment for development and review of pharmaceutical products in the USA. In response to the public uproar, in 1962 Congress enacted the Kefauver-Harris amendments to the Federal Food, Drug and Cosmetic Act. Thanks to these new amendments, manufacturers had to prove that a drug was not only safe, but also effective. Approvals had to be based on sound science. Companies had to monitor safety reports that emerged post-market and adhere to good manufacturing practices that would lead to consistently safe products including new protections for patients. This resulted in an intensive increase in animal testing across a broad range of species in varying stages of pregnancy and life cycle [8].

The new authorities given to FDA by the Kefauver–Harris Amendments were that:

- Manufacturers required to prove the effectiveness of drug products before they go on the market, and afterwards report any serious side effects.
- Evidence of effectiveness should be based on adequate and well-controlled clinical studies conducted by qualified experts. Study subjects would be required to give their informed consent.
- FDA given 180 days to approve a new drug application, FDA approval required before the drug could be marketed in the USA.
- FDA mandated to conduct a retrospective evaluation of the effectiveness of drugs approved for safety—but not for effectiveness—between 1938 and 1962.
- FDA given authority to set good manufacturing practices for industry and mandated regular inspections of production facilities.

- FDA given control of prescription drug advertising, which would have to include accurate information about side effects.
- Controlled marketing of generic drugs to keep them from being sold as expensive medications under new trade names

The genesis of subsequent regulatory requirements by FDA (and eventually other regulatory authorities) for assessing the possible reproductive and developmental effects in humans can be directly attributed to Thalidomide and the public outrage which occurred from this tragedy. As required by the Kefauver-Harris Amendments, FDA promulgated "Guidelines for Reproductive Studies for Safety Evaluation of Drugs for Human Use" in 1966 [9].

These guidelines encompassed three test intervals: Phase I (Segment I), with pre-breeding and mating exposures, to provide information on possible effects on breeding, fertility, implantation, preimplantation and postimplantation development; Phase II (Segment II), with exposures during major organogenesis, to provide information on possible effects on in utero survival and morphological growth and development, including teratogenesis; and Phase III (Segment III), with exposures from the onset of the fetal period through weaning of the offspring, to provide information on parturition, lactation, late intrauterine (fetal) and postnatal growth and development, puberty, and reproductive function of F1 offspring [10, 11].

Following the 1966 Guidelines, [12] other countries promulgated laws, regulations and guidelines for reporting and evaluating the data on safety, quality and efficacy of new medicinal products.

- Japanese Guidelines of Toxicity Studies, Notification No. 118 of the Pharmaceutical Affairs Bureau, Ministry of Health and Welfare. 2. Studies of the Effects of Drugs on Reproduction, Yakagyo Jiho Co., Ltd., Tokyo, Japan, 1984 [13].
- Canada Ministry of Health and Welfare, Health Protection Branch, The testing of chemicals for carcinogenicity, mutagenicity and teratogenicity. The Ministry of Ottawa, 1973 [14].
- UK, Committee on Safety of Medicines: Notes for guidance on reproduction studies, Department of Health and Social Security, Great Britain, 1974 [15].
- Organization for Economic Cooperation and Development (OECD), Guideline for Testing of Chemicals: Teratogenicity, Director of Information, Paris, France, 1981 [16].

Concurrently there was a realization of a divergence of regulations in the global market which created duplication of nonclinical safety studies, added expenses, and delayed development time to all potential pharmaceuticals products.

1.3 ICH Establishment and Guidelines

The need for harmonization of the regulatory requirements for pharmaceuticals was recognized after the proliferation of regulations and guidelines for reporting and evaluating safety data. By the 1980s, the European Union moved towards the development of a single market for pharmaceuticals. At the same time there were bilateral discussions between Europe, Japan, and the USA on possibilities for harmonizing regulations and with input from the WHO International Conference of Drug Regulatory Authorities (ICDRA), in Paris, in 1989 a specific plan was instituted. The birth of ICH took place at a meeting in April 1990, hosted by European Federation of Pharmaceutical Industries and Associations (EFPIA) in Brussels. Representatives of the regulatory agencies and industry associations of Europe, Japan, and the USA met, primarily, to plan an International Conference but the meeting also discussed the wider implications and terms of reference of ICH. At the first ICH Steering Committee (SC) meeting of ICH the Terms of Reference were agreed and it was decided that the Topics selected for harmonization would be divided into Safety, Quality and Efficacy to reflect the three criteria which are the basis for approving and authorizing new medicinal products [17].

One of the first Guidelines agreed to and finalized by ICH was The ICH Harmonised Tripartite Guideline Detection of toxicity To Reproduction For Medicinal Products and Toxicity *S5(R2) Detection of Toxicity to Reproduction for Medicinal Products & Toxicity to Male Fertility which was finalized in 1993 with the addendum for male fertility being finalized in 1995* [18]. The stated intent of S5(R3) is "to consolidate a strategy based on study designs currently in use for testing of medicinal products; it should encourage the full assessment on the safety of chemicals on the development of the offspring. It is perceived that tests in which animals are treated during defined stages of reproduction better reflect human exposure to medicinal products and allow more specific identification of stages at risk. While this approach may be useful for most medicines, long term exposure to low doses does occur and may be represented better by a one or two generation study approach.

The actual testing strategy should be determined by:

- anticipated drug use especially in relation to reproduction,
- the form of the substance and route(s) of administration intended for humans
- making use of any existing data on toxicity, pharmacodynamics, kinetics, and similarity to other compounds in structure/activity."

The guideline is designed to allow flexibility for testing strategies and is intended to allow scientists to design studies to best evaluate safety in the animal studies to the intended use of the drug in humans [18].

1.4 Reproductive Life Cycle and ICH Approach

The reproductive life cycle of humans is complex and a combination of studies will be required to allow exposure of mature adults and all stages of development from conception to sexual maturity. To allow detection of immediate and latent effects of exposure, observations should be continued through one complete life cycle, i.e., from conception in one generation through conception in the following generation. For convenience of testing this integrated sequence can be subdivided into the following stages:

A. Premating to conception (adult male and female reproductive functions, development and maturation of gametes, mating behavior, fertilization).

B. Conception to implantation (adult female reproductive functions, preimplantation development, implantation).

C. Implantation to closure of the hard palate (adult female reproductive functions, embryonic development, major organ formation).

D. Closure of the hard palate to the end of pregnancy (adult female reproductive functions, fetal development and growth, organ development and growth).

E. Birth to weaning (adult female reproductive functions, neonate adaptation to extrauterine life, preweaning development and growth).

F. Weaning to sexual maturity (postweaning development and growth, adaptation to independent life, attainment of full sexual function).

Stages A and B are assessed in fertility and early embryonic development to implantation study (ICH 4.1.1). An embryo–fetal development (EFD) study encompasses stage C, whereas the assessment of prenatal and postnatal development (ICH 4.1.2) comprises Stage D, F. In Nonhuman Primates (NHPs), the potential adverse effects of drugs of a drug product on different stages of reproductive cycle defined as stages A–F cannot be fully assessed as in rodents and rabbit. Full assessment of stage A is not feasible, stage B cannot be easily evaluated because the earliest dosing of pregnant monkey is on GD 20 with the implantation window of cynomolgus monkeys ranging between GD 9 to GD 15. Stage C is assessed in NHP DART study. The enhanced prenatal and postnatal development (ePPND) study in NHP involves assessment of stage D, E of reproductive cycle, but rather than the examination of the neonate being extended to weaning, it limits the evaluation of adaptation of the neonate to extrauterine life and may include preweaning development and growth [19].

The reproductive cycle while defined in segments above is continuous and therefore a toxic insult can occur in one part of the cycle while having consequences in another part of the

cycle. An example would be exposure of gametes (either sperm or eggs) to a toxic insult which would not be observed until later in the reproductive cycle, i.e., an affect on mating resulting in reduced fertility. The ICH study designs allow evaluation of both the male and female components of mating and fertility. Effects can also occur across generations depending on the timing of exposure. One of the critical findings from studies conducted on Thalidomide was that the timing of exposure as to gestational age of the fetus is critical for the malformations which were observed. Other tragedies such as children exposed to methyl mercury (Minamata Disease) as a result of maternal ingestion resulted in severe impairment of central nervous system development including: mental retardation, cerebellar ataxia, physical growth disorder, dysarthria, and limb deformities. Most of them showed hyperkinesis, hypersalivation, seizures, and strabismus [20] While there were only trivial or no symptoms of intoxication in mothers, infants showed central nervous disturbances such as paralysis and intelligence disorders [21]. This tragedy brought to light the need to consider late gestational and postpartum exposures as part of a complete nonclinical safety package. Pediatric and juvenile toxicology studies are now routinely conducted as part of a nonclinical safety package if exposures during the critical periods of neonatal and juvenile development are expected.

1.5 ICH Stages of Development

ICH Stages of Development are delineated in six stages and defined accordingly:

A. *ICH Stage A*: Premating to conception. This stage evaluates reproductive functions in adult males and females, (development and maturation of gametes, mating behavior, fertilization). This stage evaluates reproductive function in mature adults (male and female). The timing of pregnancy in this stage has been defined as the time spermatozoa are first identified either by vaginal smear or observation of a vaginal plug in rodents.

B. *ICH Stage B*: *Conception to implantation* (adult female reproductive functions, preimplantation development, implantation).

   Stage B examines reproductive function in females preimplantation and implantation for the conceptus. This period in rodents and rabbits is gestation day 6 or 7 of pregnancy depending on species.

   Stages A and B are evaluated in the study design for the Fertility and Early Embryonic Development to Implantation design as outlined in ICH 4.1.1. This design corresponds to what was formerly known as the *Segment I Study.* ICH Stages A-B are graphically represented in Fig. 1.

C. *ICH Stage C*: *Implantation to closure of the hard palate* (adult female reproductive functions, embryonic development, major organ formation). Stage C evaluates toxicity to the female

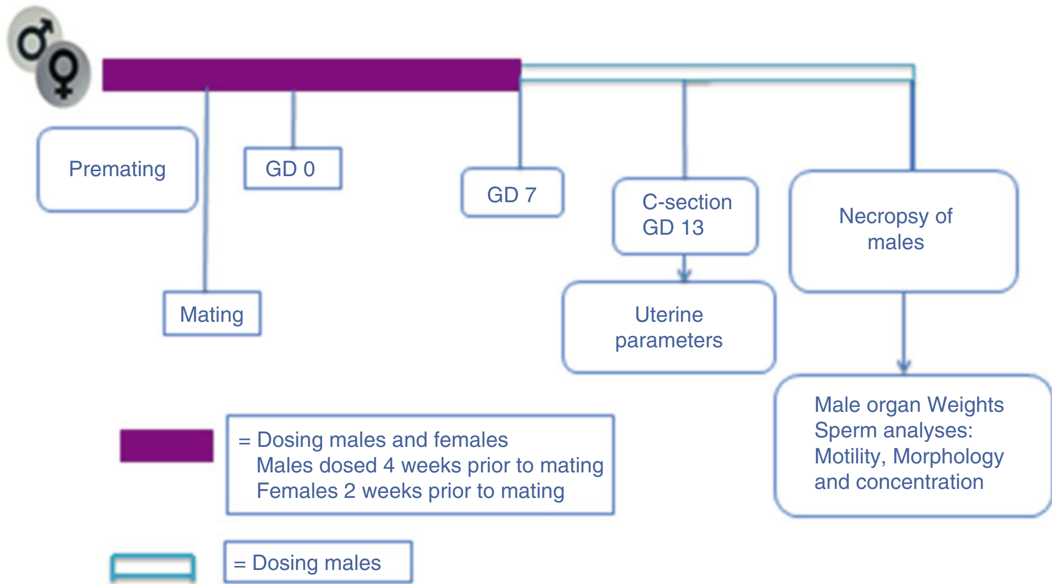

Fig. 1 ICH 4.1.1: Fertility study in rats

reproductive function from implantation through embryonic development and the period of major organogenesis in the developing fetus. The period of embryogenesis is completed with the closure of the hard palate which varies slightly as to gestational age depending on species used.

D. *ICH Stage D*: *Closure of the hard palate to the end of pregnancy* (adult female reproductive functions, fetal development and growth, organ development and growth). Stages C and D are evaluated in the Embryo–Fetal Toxicity Study which covers the developmental period from implantation, organogenesis, closure of the hard palate through fetal growth. ICH Stages C–D are graphically represented in Fig. 2.

E. *ICH Stage E*: *Birth to weaning* (*adult female reproductive functions*, *neonate adaptation to extrauterine life*, *preweaning development and growth*). This stage focuses on female reproductive functions including parturition, maternal care postpartum, lactation and nursing behavior, and overall litter care and survival of F1 pups.

F. *ICH Stage F*: *Weaning to sexual maturity.* Typically Stage F is evaluated when a pharmaceutical product is intended for use in children. This ICH stage provides observations of postweaning development and growth, (adaptation to independent life, attainment of full sexual function). ICH Stages E–F are graphically represented in Fig. 3.

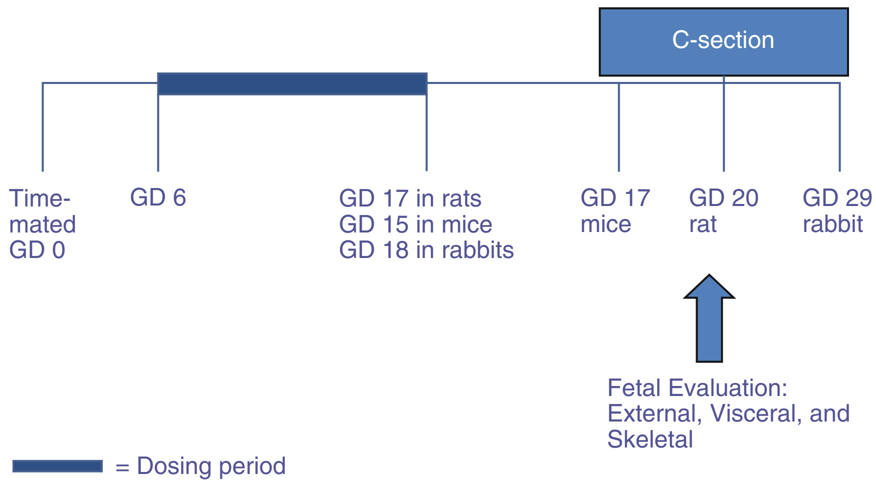

Fig. 2 ICH 4.1.3: Embryo–fetal development (segment II)

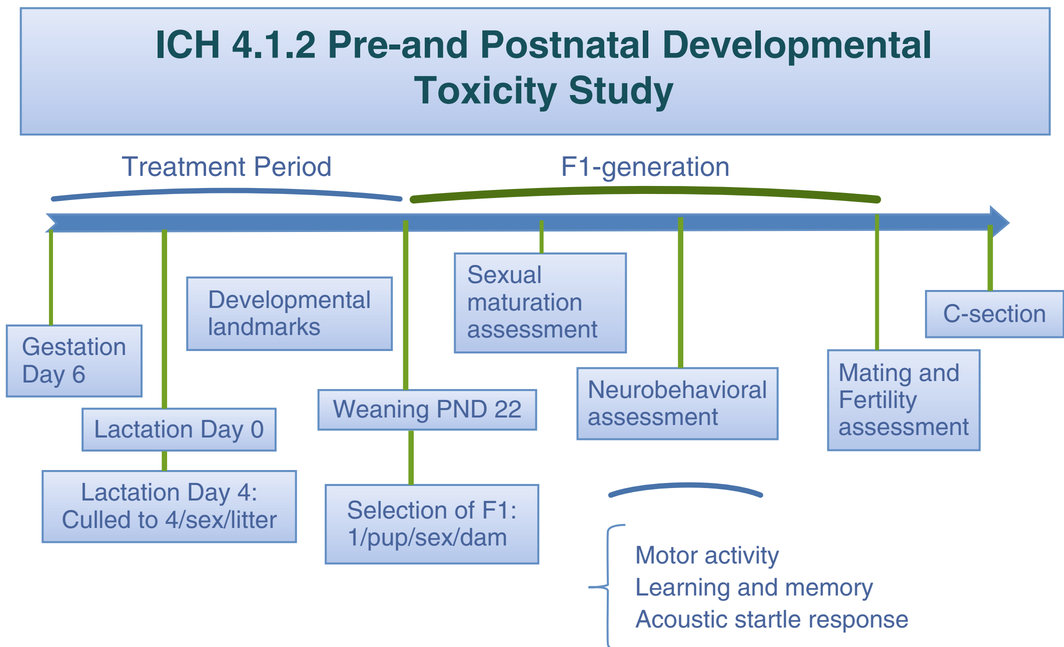

Fig. 3 ICH 4.1.2: Prenatal and postnatal developmental toxicity study

The ICH guidelines recommend:

1. Studies should be planned according to the "state of the art," and take into account preexisting knowledge of class-related effects on reproduction.
2. Scientific justification and flexible study designs to meet the needs of the use of the drug;
3. Kinetics to support interpretation of results;
4. Expanded male reproductive toxicity evaluations
5. Mechanistic studies if possible;
6. Consideration of all endpoints of developmental toxicity (embryo–fetal loss, malformations, reduced weights, and functional/behavioral deficits).

2 Current ICH Study Designs

2.1 ICH Study of Fertility and Early Embryonic Development to Implantation (ICH 4.1.1)

Evaluates the toxic effects/disturbances resulting from treatment from before mating (males/females) through mating and implantation. This comprises evaluation of stages A and B of the reproductive process (*See* Fig. 1). For females, this should detect effects on the estrous cycle, tubal transport, implantation, and development of preimplantation stages of the embryo. For males it will permit detection of functional effects (e.g., on libido, epididymal sperm maturation) that may not be detected by histological examinations of the male reproductive organs. Species selection is discussed in another section of this chapter; however, the rodent is typically utilized for the Fertility and Early Embryonic Development to Implantation Study due to the extensive historical control database established and the relative ease of assessing the reproductive parameters including: maturation of gametes, mating behavior, fertility, preimplantation stages of the embryo, and implantation [22].

The pre-ICH fertility study design (Segment 1) included some major differences from the current ICH 4.1.1 study design including treatment of male rats with the test compound for 60–80 days prior to mating to cover a complete spermatogenic cycle in the rat. Female rats were treated for 14 days (at least three estrous cycles), then mated to the treated males. Dosing was continued throughout pregnancy, with half of the females euthanized on day 13 post-coitus (pc) and examined for number and distribution of live and dead embryos (the day of confirmation of mating is designated as Gestation Day [GD] 0). The remaining dams were allowed to litter and the newborn pups were counted, examined, sexed, and weighed. Pups were counted and weighed again on postnatal days 4 and 21. A major change is the reduction in the premating dose period for males which was shortened from 70 to 28 days (4 weeks) premating. This change was made because compounds inducing

selective effects on male reproduction are rare; and there is no conclusive example of a male reproductive toxicant where dosing for 9–10 weeks and mating them with females provides data that could not be obtained with the shorter dose period outlined in the ICH Fertility study. Data on the potential effects on spermatogenesis is best obtained in repeated dose toxicity studies which include histopathologic evaluation of the testis. This allows the ICH fertility study to focus on other, more immediate toxicity to the male and female reproductive systems. In the rodent the full sequence of spermatogenesis (including sperm maturation) lasts 63 days. If there is some indication of histopathologic changes in the repeat-dose studies the dose period for males can be adjusted to cover a full spermatogenic cycle [23].

The design of the fertility study, especially the reduction in the premating dosing period for males, is based on evidence accumulated and re-appraisal of the basic research on the process of spermatogenesis that originally prompted the demand for a prolonged premating treatment period. Compounds inducing selective effects on male reproduction are rare; mating with females is an insensitive means of detecting effects on spermatogenesis; good pathological and histopathological examination (e.g., by employing Bouin's fixation, paraffin embedding, transverse sections of 2–4 μm for testes, longitudinal sections for epididymides, PAS and hematoxylin staining) of the male reproductive organs provides a more sensitive and quicker means of detecting effects on spermatogenesis; compounds affecting spermatogenesis almost invariably affect postmeiotic stages; there is no conclusive example of a male reproductive toxicant, the effects of which could be detected only by dosing males for 9–10 weeks and mating them with females [22]. The standard duration of the ICH fertility study in the rat is approximately 7 weeks, which includes 4 weeks of treatment for the males and 2 weeks of treatment for the females prior to mating, mating and the females are examined on day 13 of gestation (Day 0 determined by evidence of mating). However, if repeated dose toxicity studies (general toxicology studies) show there are effects on weight or histology of reproductive organs in male or female rats, the need for a more comprehensive study should be considered. Among the possibilities are a male fertility study with an extended administration period, and a mating period after 10 weeks of dosing (the time it takes for a spermatogonium to become a functional mature sperm), rather than 4 weeks of premating dosing. In the event effects on fertility are expected (as they would be with certain classes of drugs, e.g., contraceptives, oncology drugs, steroids) a recovery period should be included in the study design.

2.2 Species Selection

As stated in the ICH Guideline the fertility study should be conducted in a mammalian species. It is generally desirable to use the same species and strain as in other toxicological studies. There are

scientific justifications for this as the early toxicology studies establish appropriate dose regimens and toxicokinetic information that can directly apply to the design of the study. Reasons for using rats as the predominant rodent species are practicality, comparability with other results obtained in this species and the large amount of background knowledge accumulated. For some compounds the most appropriate species may be another rodent such as the mouse. Other species may be appropriate but the reproductive cycle of the particular species must be taken into account when designing the study.

Each species has advantages and disadvantages. Rats and to a lesser extent mice, are good, general purpose models; the rabbit has been somewhat neglected as a "non-rodent" species for repeated dose toxicity and other reproduction studies than embryotoxicity testing. It has attributes that would make it a useful model for fertility studies, especially male fertility. For both rabbits and dogs (which are often used as a second species for chronic toxicity studies) it is feasible to obtain semen samples without resorting to painful techniques (electro-ejaculation) for longitudinal semen analysis. Most of the other species are not good, general purpose models and probably are best used for very specific investigations only. The general principle here as with other toxicology studies is to select the most appropriate species based on the evidence that the compound being tested is pharmacologically/toxicologically active in the species selected.

All species have their disadvantages, for example:

1. *Rats*: sensitivity to sexual hormones, unsuitable for dopamine agonists due to dependence on prolactin as the primary hormone for establishment and maintenance of early pregnancy, highly susceptible to nonsteroidal anti-inflammatory drugs in late pregnancy.
2. *Mice*: fast metabolic rate, stress sensitivity, malformation clusters (which occur in all species) particularly evident, small fetus.
3. *Rabbits*: often lack of kinetic and toxicity data, susceptibility to some antibiotics and to disturbance of the alimentary tract, clinical signs can be difficult to interpret.
4. *Guinea pigs*: often lack of kinetic and toxicity data, susceptibility to some antibiotics and to disturbance of the alimentary tract, long fetal period, insufficient historical background data.
5. *Domestic and/or mini pigs*: malformation clusters with variable background rate, large amounts of compound required, large housing necessary, and insufficient historical background data.
6. *Ferrets*: seasonal breeder unless special management systems used (success highly dependent on human/animal interaction), insufficient historical background data.

7. *Hamsters*: intravenous route difficult if not impossible, can hide doses in the cheek pouches and can be very aggressive, sensitive to intestinal disturbance, overly sensitive teratogenic response to many chemicals, small fetus.
8. *Dogs*: seasonal breeders, inbreeding factors, insufficient historical background data.
9. *Nonhuman primates*: kinetically they can differ from humans as much as other species, insufficient historical background data, often numbers too low for detection of risk. They are best used when the objective of the study is to characterize a relatively certain reproductive toxicant, rather than detect a hazard [18]

2.3 Considerations for Alterations in Study Design

There are iterations on the standard fertility protocol if indications from previous toxicology studies indicated there may be a sex difference in the toxicology profile of the compound. In this case, untreated males can be mated to treated females with these males then being treated and mated to untreated females. Separate effects can then be evaluated by this selective dose design.

Another alternate design is to increase the number of animals in the control and high dose groups and employ a cross mating design between the two groups. This design can also be employed to evaluate sex related differences in fertility. In design, the study would initiate with the standard 2-week treatment period for the females which are then mated to untreated males. These males would then initiate their treatment period and then be mated to untreated females.

While not commonly employed, the female treatment period can be extended to gestation day 17 or 19 with a caesarian section exam on gestation day 21. There is overlap of course with the embryo–fetal developmental toxicity study but in some cases this is a viable option. This design combines assessment of male and female fertility, reproductive performance and developmental toxicity.

Extension of the female dose period is also possible from gestation and then through parturition to lactation Day 21. As components of this design are evaluated in other study designs this scenario is not commonly employed.

2.4 Dose Selection

The main objective of regulatory toxicology studies is to establish the potential hazards associated with the test item by identifying potential organ toxicity. For the fertility study design dose selection is critical to establishing toxicity to the male and female reproductive systems. The primary parameter used in dose selection is the tolerability of the test item in animals. Tolerability is determined during toxicity testing in single and multiple dose studies conducted prior to initiation of the fertility study by observations such as clinical signs, reductions in body weight or changes in

food consumption. Key parameters such as systemic exposure and histopathology may also be used to support dose selection. The clinical condition of an animal usually gives an initial indication that the test item is causing systemic toxicity. The current Committee for Proprietary Medicinal Products (CPMP) note for guidance on repeated dose toxicity studies indicates that doses should be selected to establish a dose or exposure response to treatment. Most study designs incorporate the use of three dose groups of animals receiving the test item, at low, intermediate and high doses, plus a control group which receives vehicle alone. Regulatory authorities expect the high dose selected for regulatory nonclinical toxicology studies to produce evidence of toxicity [24]. The CPMP guidance also indicates that the high dose should be selected to enable identification of target organ toxicity, or other nonspecific toxicity, or until limited by volume or limit dose [25].

There are five general criteria for defining the high dose in a toxicology study. These are (1) maximum tolerated dose, (2) limit dose, (3) top dose based on saturation of exposure, (4) maximum feasible/practical dose, or (5) dose providing a 50-fold margin of exposure. A more complete discussion of dose selection can be found in a description of the options for selecting the high dose in general toxicity studies see ICH guidance M3 (R2) [25].

The fertility study is typically conducted following the IND enabling studies which depending on the test article will include multiple dose toxicity study data which can be used to set doses for the fertility study. Often the 90-day study will be completed and histopathologic data from the males will be used to determine the most appropriate dose period for the males in the fertility study. If only data from the 28-day toxicity study are available, a longer dose period for the males should be considered particularly if there is evidence of effects on reproductive organ weights. This is applicable for both males and females and should be utilized in the design of the fertility study as discussed above in Considerations for Alterations in Study Design.

Determination of the high dose should use data obtained from repeat dose toxicity studies or from preliminary reproduction studies. Parameters such as: reduction in bodyweight gain, increased bodyweight gain, particularly when related to perturbation of homeostatic mechanisms, specific target organ toxicity, hematology, clinical chemistry, exaggerated pharmacological response, which may or may not be reflected as marked clinical reactions (e.g., sedation, convulsions). Additional considerations should include the physicochemical properties of the test substance or dosage formulation which, allied to the route of administration, may impose practical limitations in the amount that can be administered. Under most circumstances 1 g/kg/day should be an adequate limit dose. Pharmacokinetic parameters should also be assessed as they are useful in determining high dose exposure for

low toxicity compounds. There is, however, little point in increasing administered dosage if it does not result in increased plasma or tissue concentration. Once the high dosage has been set, lower dosages should be selected in a descending sequence, the intervals depending on kinetic and other toxicity factors. If at all possible, it is recommended a "no observed adverse effect level," should be given priority when setting the low and mid dosages. The selection of the mid dose is typically chosen as an interval between the low and high doses, however, its selection should include dosage intervals close enough to reveal any dosage related trends that may be present.

2.5 Number of Animals and General Study Design Considerations

Most study designs for the fertility study include 20–25 animals per sex per group. The number of animals per group is larger than the number typically included in the standard repeat dose toxicity tests. This is done to increase the power of the tests and assessments conducted in the fertility study such as sperm evaluations, estrus cycling, and mating behavior as these evaluations are typically only done in this study whereas other parameters such as body weights and histopathology are repeatedly evaluated in the standard toxicity tests.

3 Fertility and Reproduction Evaluation

3.1 Female Fertility Evaluation

3.1.1 Parameters

Parameters that should be included in the evaluations of Fertility and Early Embryonic Development include those conducted in general toxicity studies: body weights, food consumption, clinical observations, and toxicokinetic evaluations. In addition, the fertility study also evaluates female reproductive parameters including: estrus cycling evaluations, mating behavior, fertility assessments, ovarian (weight and number of corpora lutea) and uterine examinations which include implantation counts, and an assessment of viability of implantations. Organ weights and histopathologic evaluations of the reproductive organs for both males and females are required to fully examine the effects of the test article on fertility and reproduction. For the males, evaluation of sperm for total count, concentration, motility, and morphology add to the overall assessment of reproductive function.

Estrus Cycling Evaluation is relatively easy to conduct in the commonly used rodent species (rat and mouse) by a vaginal swab or lavage followed by visual assessment of the cells in the wash by trained technicians. In rodents, there are four stages defined by cell type: proestrus, estrus, metestrus, and diestrus which repeat every 4–5 days unless interrupted by pregnancy, pseudopregnancy, or anestrus. While the estrus cycle is defined by stages determined by the hormonal status of the female it is important to understand there is a continuum and visual assessment should be conducted by

experienced evaluators. When the female is in proestrus, mostly nucleated and some cornified epithelial cells are present. Some leukocytes may be present if the female is in early proestrus. As the stage of the cycle advances to estrus, mostly cornified epithelial cells are present. If the cycle is not interrupted by pregnancy, pseudopregnancy, or other phenomena, metestrus will begin. Metestrus is a brief stage when the corpora lutea form but fail to fully luteinize due to a lack of progesterone. The uterine lining will begin to slough and evidence of this is seen in the form of cornified epithelial cells and polymorphonuclear leukocytes present in vaginal swabs. Some nucleated epithelia cells will also be present in late metestrus. Diestrus is the longest of the stages lasting more than 2 days. Vaginal swabs during diestrus show primarily polymorphonuclear leukocytes and a few epithelial cells during late diestrus. Leukocytes remain the predominant cell type having removed cellular debris. The cycle then repeats [26]. In the fertility study estrus cycling is evaluated during the 14 day pre-dose period the 14 day and continues through cohabitation and until confirmation of mating. Abnormally cycling females in the pre-dose period should be excluded from the study prior to dose administration. If hamsters are used the evaluation of estrus cycling is similar but should also include visual observation of the vaginal opening and perivaginal area for evidence of vaginal discharge. In the event the guinea pig must be used (which is rare) assessment of estrus (cycle is up to 18 days) is not practical due to the closure of the vaginal opening by a membrane.

For rodents mating should be done on a 1:1 ratio to allow for easy evaluation of an effect on fertility. Mating/copulation in rats, mice, and hamsters is confirmed by presence of sperm in the vaginal smear or presence of a copulatory plug in situ. For rodents mating performance is commonly measured by the following endpoints:

1. Number of days in cohabitation
2. Number of females confirmed mated
3. Pregnancy index calculated by total population per group (% pregnant/number females cohabited) and inseminated population per group (% pregnant/number inseminated).

In the event a female does not mate with the paired male during the first cohabitation period, she may be paired with another male and the period of the second cohabitation should be recorded along with identification of the second male and whether mating occurs (or not). There are a number of confounding factors that can affect mating behaviors particularly in mice such as environmental conditions or changes in housing. For rats mating behavior can also be affected by changes in photoperiod, nutrition, stress, and compounds that affect the endocrine system.

3.1.2 Evaluation of Mating Success: Preimplantation Loss, Implantation Loss, and Implantation Viability

Mated females are euthanized and evaluated typically on gestation day 13 close to the beginning of the fetal period and a laparotomy is conducted to examine the number and status of implantation sites. At necropsy the ovaries are examined for corpora lutea and the total number of corpora lutea are compared to the number of total implantations.

Preimplantation loss: (number of implantation site − number of fetuses/number of implantation sites × 100)

Postimplantation loss: (number of implantation sites − number of fetuses/number of implantation sites × 100)

Implantation Viability: (number of live fetuses/number of fetal deaths).

3.1.3 Ovarian Evaluation and Follicular Count

An optional assessment for evaluating potential effects of the female reproductive function is via evaluation of ovarian toxicity by comprehensive histopathological examination of the female reproductive organs based on the underlying morphology of a normal cycle of the reproductive tract including the ovary to identify pathological findings of ovarian toxicity (decreases in follicles, increases in atretic follicles, increases in currently formed corpora lutea, etc.) can be a useful toll to evaluate effects reflected in the female fertility parameters (irregular estrous cycle, preimplantation loss).

Accurate estimation of the number of ovarian follicles at various stages of development is an important indicator of the process of folliculogenesis in relation to the endocrine signals and paracrine/autocrine mechanisms that control the growth and maturation of the oocytes and their supporting follicular cells [27]. Ovarian histopathology can be a useful assessment tool in the fertility study to evaluate the development of follicles within the ovary. Ovaries are under the tight control of both hormones and growth factors. Stimulation of follicular growth and the accompanying developmental arrest of most follicles lead to the cyclical nature of mammalian female reproduction, and ultimately the release of viable oocytes [28]. The complement of developing follicles within the ovary originates from and is dependent on the immature nongrowing stock of primordial follicles. Coordinated entry of these follicles into the growth phase controls the rate at which the follicular reserve is depleted [28–31]. Evaluation of the ovary is an important endpoint in toxicological assessments since the loss of oogonia, oocytes, or supportive somatic cells may have adverse effects on reproduction. Properly conducted follicle counts can supplement qualitative ovarian assessment to characterize ovarian toxicants, understand their site of action, and assess primordial follicle integrity when ovarian lesions are subtle. However it should be recognized that data generated from follicle counts can be highly variable, making interpretation difficult. When depletion or lesions of small follicles are clearly identified during a qualitative histologic examination, follicle counts may add little or no additional

information for chemical hazard identification. The evaluation of ovarian toxicity should use a weight-of-evidence approach, considering the qualitative histopathology data and other available data (fertility data, clinical observations, organ weights, male reproductive toxicity data, etc.). The examination should be conducted by a toxicologic pathologist familiar with the normal reproductive cycle in the species, and should include evaluation of all major components of the ovary (follicles, corpora lutea, stroma, interstitium, and vasculature), with special attention given to the qualitative assessment of primordial and primary follicles. This qualitative assessment of ovaries should be done in conjunction with microscopic evaluation of the entire reproductive tract, and with consideration of all ancillary reproductive data (organ weights, estrous cyclicity, etc.) [32].

3.2 Male Fertility Evaluation

In ICH S5 A and B Detection of Toxicity to Reproduction for Medicinal Products & Toxicity to Male Fertility states that: "information on potential effects on spermatogenesis can be derived from repeated dose toxicity studies" [18]. The original ICH S5 guideline (1993) was amended in 1996 with Part II: Toxicity to Male Fertility which incorporated into the core guideline that sperm analysis (sperm counts, sperm motility, sperm morphology) is optional and may be used to confirm findings by other methods or to characterize effects further. Sperm analysis may be added to the study design if general toxicology studies suggest an effect on the testicular epithelium or other reproductive organs, or if fertility is noted to be affected in an ongoing male fertility study. Unlike humans laboratory species such as rodents are likely better able to sustain reductions in sperm count without affecting fertility, that is, they have a higher "reserve capacity." Therefore, relative effects on sperm parameters may be of value, in cases an, extrapolation of risk to man is desired [23].

Sperm analysis data are considered more relevant for fertility assessment when samples from vas deferens or from cauda epididymis are used. These samples can only be collected at necropsy in rodents, and all relevant findings to make the decision to collect such samples may not be known at this time. Analysis of sperm motility requires fresh samples, while sperm counts and morphology can be obtained from fixed or frozen samples. In addition, analysis of sperm motility most often relies on use of CASA (computer assisted sperm analysis) systems. Some laboratories will routinely collect and analyze samples for some or all of these parameters [23].

The full sequence of spermatogenesis (including sperm maturation) in rats lasts 63 days. When the available evidence, or lack of it, suggests that the scope of investigations in the fertility study should be increased, or extended from detection to characterization, appropriate studies should be designed to further characterize the effects [18]. Inclusion of measures of rat sperm motility and

morphology, into the fertility protocol can increase the information obtained from the study particularly in the event an effect on fertility is observed. Computer-aided sperm analysis allows assessment of sperm counts, motility, and morphology, allowing an extensive evaluation of rat spermatozoa. This technology can provide an objective means of classifying the motion of rat spermatozoa as progressive or non-progressive, as required in test protocols. More specific tests of rat sperm function are being applied for the purpose of evaluating modes and mechanisms of toxicant action [23].

To assess effects on the male in fertility studies, endpoints investigated should include the most sensitive parameters, i.e., histopathology of the testes. Other sensitive indicators of toxicity include the weights of reproductive organs and the accessory glands (testis, epididymis, prostate, and seminal vesicle) as well as sperm parameters such as sperm count, sperm morphology, and sperm motility. Sperm motility was found to be in some cases more sensitive than histopathology. The above parameters showed a higher sensitivity than fertility parameters [33]. Sperm assessments are commonly conducted now using Computer Assisted Sperm Analysis (CASA) technology which can provide rapid assessment of sperm count, sperm morphology, and sperm motility. Further methods have been described by Seed et al. [34].

Each species that may be used for assessment of fertility have differing lengths of spermatogenesis and the particular anatomical and physiologic peculiarities of each model should be taken into account when evaluating histopathologic and physiologic effects on the male reproductive system. In rare cases, staging of the testis/seminiferous tubules may be needed to determine targeted effects on the male; however, due to sperm production in the most commonly utilized species (rats and mice) an effect on sperm production sufficient to affect fertility is rare.

3.2.1 Hormone Regulation of Fertility

In both female and male mammals, gametogenic and endocrine functions of the reproductive organs are regulated primarily by two pituitary gonadotropins, luteinizing hormone (LH) and follicle-stimulating hormone (FSH). In addition, prolactin (PRL) is involved in the control of the corpus luteum in many species. It should be noted there is on significant difference between rodents and humans in hormonal regulation of reproduction. Prolactin in rodents acts in conjunction with LH and FSH on luteal function in rodents but does not play a similar role in humans. Ovarian function can also be influenced by growth hormone (GH), directly or via stimulation of systemic and/or local production of the insulin-like growth factor (IGF) IGF-I, and that testicular function can be influenced by GH, IGF-I, and PRL. Studies of PRL, GH, and IGF-I binding and of the expression of the corresponding receptors have identified numerous potential targets for the action of these hormones on reproductive processes [35]. These targets include

various neuronal groups in the hypothalamus, the anterior pituitary, and different types of somatic cells in the gonads as well as both the male and the female reproductive tracts [36]. The regulation of reproduction is complex and interruption of hormonal control of reproduction can result in effects on fertility as readily as direct insult to reproductive cells, tissues and organs. Reproduction in mammals is controlled directly and indirectly by the hypothalamic–pituitary–gonadal axis that coordinates reproductive behavior with ovulation. The primary signal from the central nervous system is gonadotropin-releasing hormone (GnRH), which modulates the activity of anterior pituitary gonadotropes regulating follicle stimulating hormone (FSH) and luteinizing hormone (LH) release. As ovarian follicles develop they release estradiol, which negatively regulates further release of GnRH and FSH. As estradiol concentrations peak they trigger the surge release of GnRH, which leads to LH release inducing ovulation. Release of GnRH within the central nervous system helps modulate reproductive behaviors providing a node at which control of reproduction is regulated. This requires the coordination of peripheral organs with the nervous system to ensure that the internal and external environments are optimal for successful procreation of the species [36].

4 Embryo–Fetal Developmental Toxicity Study Designs

4.1 General Background

Evaluation of the hazards posed by potential new pharmaceutical products to the developing fetus is in many respects the centerpiece of safety evaluations outlined in ICH S5(R2) and in effect the embryo–fetal development studies were a direct regulatory response to the thalidomide tragedy. Identification of chemicals or biologics that have the potential to affect the developing fetus is a primary concern for both the pharmaceutical industry and regulatory authorities alike. Assessment of potential developmental and reproductive toxicity of human pharmaceuticals is currently guided by the International Conference on Harmonisation ICH S5(R2). The studies that assess developmental hazard are generally conducted in rodents and rabbits. Regulatory agencies rely on data obtained from this study protocol to ascertain the potential risks to human development after prenatal exposure to a particular chemical. There is strong scientific rationale for using data from animal studies due to the conservative nature of developmental processes. However, it is also recognized that the potential for species-specific differences exists; consequently, some kinds of effects observed in animal prenatal developmental toxicity studies may not be particularly predictive of the potential for adverse effects in humans. An understanding of what species differences exist and their potential impact on the response to chemical insult is crucial to improving our ability to accurately predict risks to humans.

4.2 Wilson's Six Principles of Teratology

1. Susceptibility to teratogenesis depends on the genotype of the conceptus and the manner in which this interacts with environmental factors.
2. Susceptibility to teratogenic agents varies with the developmental stage at the time of exposure.
3. Teratogenic agents act in specific ways (mechanisms) on developing cells and tissues to initiate abnormal embryogenesis (pathogenesis).
4. The final manifestations of abnormal development are death, malformation, growth retardation, and functional disorder.
5. The access of adverse environmental influences to developing tissues depends on the nature of the influences (agent).
6. Manifestations of deviant development increase in degree as dosage increases from the no-effect to the totally lethal level [37].

Wilson also proposed mechanisms of developmental toxicity which included: mutation, chromosome nondisjunction and breaks, mitotic interference, altered nucleic acid integrity or function, lack of precursors or substrates for biosynthesis, altered energy sources, enzyme inhibitions, osmolar imbalance, and altered membrane characteristics [37].

While Wilson's mechanisms are still salient we now have added tools to understand the molecular mechanisms that have been intensively studied over the past few years and are now known to be important in embryonic development. These include the following:

- Epigenetic control of gene expression
- The effects of small regulatory RNAs
- The imbalance of gene products resulting from submicroscopic alterations of genomic structure such as copy number changes
- Alterations of the cytoskeleton
- Perturbations of the extracellular matrix
- Effects of mechanical forces on embryogenesis
- Disturbances of intracellular or intercellular signaling
- Dysfunction of molecular chaperones
- Effects on the distribution of molecules into subcellular compartments
- Alterations of the integrity of intracellular organelles [38]

Embryo–fetal developmental toxicology studies cover ICH Stages C and D which specifically evaluates the developing conceptus through the period up to parturition. As outlined in ICH S5 (R2), 4.1.3; two species (a rodent and a non-rodent) are commonly used for evaluation of small molecules. Evaluation of embryo–fetal

developmental toxicants has received extensive regulatory emphasis since thalidomide and for most drug development programs these studies are critical for a new chemical entity to proceed in the development process. Because of the sensitivity of the rabbit to thalidomide, this species is commonly used as the non-rodent model for the embryo–fetal development studies. As the mix of new pharmaceutical products has shifted from almost exclusively small molecules when the guidelines were written to today's environment where more large molecules and biopharmaceuticals are a significant portion of drugs entering development, consideration of the most appropriate non-rodent species must be taken into account. Developmental and reproductive toxicology testing in nonhuman primates (NHPs) has become more common due to the increasing number of biopharmaceuticals in drug development, since NHPs are frequently the only species to express pharmacologic responses similar to humans. Nonhuman primates (NHP) are most frequently used for Developmental and Reproductive Toxicity (DART) testing when rodents and/or rabbits are not the most pharmacologically relevant species. For a species to be pharmacologically relevant, the test molecule must produce similar in vitro or in vivo pharmacology in the test species and in humans [39, 40].

As noted in the ICH guidance "choosing an animal species and strain for reproductive toxicity testing care should be given to select a relevant model. Selection of the species and strain used in other toxicology studies may avoid the need for additional preliminary studies. If it can be shown—by means of kinetic, pharmacological and toxicological data—that the species selected is a relevant model for the human, a single species can be sufficient. There is little value in using a second species if it does not show the same similarities to humans. Advantages and disadvantages of species (strains) should be considered in relation to the substance to be tested, the selected study design and in the subsequent interpretation of the results." [41]

4.3 Study Design Requirements

The study designs for the embryo–fetal developmental toxicity studies are outlined in the International Conference on Harmonisation (ICH S5(R2)). This section will concentrate on the study designs for the most commonly conducted embryo–fetal development studies typically conducted for small molecules. The design of studies to support registration of biologics (e.g., vaccines, monoclonal antibodies, and blood products) will be discussed in the context of adding to the study overview, however, since there are no formal guidelines for the design of developmental toxicity studies to assess the safety of these agents. For vaccines, there are recommendations from the FDA for the conduct of developmental toxicity studies [42]. ICH does not have definitive separate guidance for large or biologic molecules there is an understanding that biopharmaceuticals (i.e., human proteins and peptides) are also governed by ICH S5(R2). However, because some human proteins

may not cross-react with the standard species used for evaluation of EFD or may be immunogenic in animals, nonstandard species or nonstandard approaches may need to be adopted for some biopharmaceuticals.

4.3.1 Species Considerations

Guidelines specify a rodent (typically the rat or mouse) but hamsters can be used if they are the most pharmacologically relevant species. The non-rodent second species is typically the rabbit, however, as noted the most relevant species for biopharmaceuticals may be the nonhuman primate (NHP). The species selected should be the most pharmacologically relevant species.

4.3.2 Conducting Dose Range-Finding Studies for the Definitive Embryo–Fetal Development Studies

Many laboratories utilize data from their repeat dose toxicity studies conducted in the relevant rodent species as part of the IND enabling package to determine dose levels for the definitive embryo–fetal development study (dEFD). While this practice is acceptable and generally does not result in unexpected toxicities in the pregnant female, there are cases where the pregnant animal may present a very different toxicology profile from the nonpregnant animal. This can be due to physiologic and hormonal differences between nonpregnant and pregnant animals. In the case of the non-rodent species, rabbits are the most commonly used second species and only rarely would any toxicology data be available in this species. Therefore, it is advisable to conduct a preliminary embryo–fetal development study (pEFD) in the rabbit.

The designs of range-finding, pEFD, and dEFD studies are similar except for the number of mated females per group and the extent of fetal examinations. As per ICH guidance, the study designs should be planned to take into account preexisting knowledge of class-related effects, should avoid suffering of animals, and use the minimum number of animals necessary to achieve overall objectives.

In most cases, information from general toxicology studies is available prior to the start of a rodent study with pregnant females; however, these data in nonpregnant females do not assure the same dose response in pregnant females. In addition, it is very rare to have any toxicological information in rabbits. Due to the possibility of increased or decreased sensitivity in pregnant animals, a dose range-finding study should be considered before the conduct of the definitive EFD studies. Another approach that may be particularly helpful for rabbits and will tend to minimize the risk of excessive toxicity is to conduct the range-finding or pEFD study in stages. The first stage may consist of two widely spaced dose levels, and depending on the tolerability of these doses, additional stages are added as needed to assess lower, intermediate, or higher dose levels. Some laboratories may conduct a range-finding study in nonpregnant rabbits first.

There is no ICH guidance for the design of range-finding studies. An adequate range-finding study would include at least six mated females per group and at least three drug-treated groups, although more animals per group and up to five drug-treated groups are common. The duration of dosing, maternal parameters, and caesarean section parameters are generally the same as dEFD. Fetal examinations can be limited to only external observations since together with resorptions and fetal weights an adequate limited, assessment of developmental toxicity can provide sufficient data to set doses for the definitive study. An alternative design for a rodent range-finding study that would also serve to select dose levels for the subsequent prenatal and postnatal study is to extend the dosing period to approximately Lactation Day 6 and allow the females to deliver naturally. Pups would be evaluated for external morphology, survival, and growth in this study design to Postnatal Day 7. Alternatively, dams could be dosed and pups followed through subsequent days of the preweaning period to obtain additional information. Although it is possible that cannibalism could obscure observations of morphological abnormalities, it is likely that other parameters would also be affected [43]. In addition, there may be significant differences in drug effects on embryo–fetal and postnatal offspring survival, which should be considered when selecting appropriate doses for the EFD and PPN studies based on this extended dose range–finding study (i.e., dose selection for the EFD study should not be based on increased peri/neonatal mortality) [43].

4.3.3 Number of Animals

ICH recommends 16–20 pregnant animals per group with four groups (three treated and one vehicle control group) at a minimum for the definitive embryo–fetal development study in rodents and for the non-rodent study 16 pregnant animals per group. In the experience of the author, most laboratories conducting these studies assign 20–22 animals per group for the rodent study and 20 animals per group in the case of the rabbit study. Pregnancy rates form vendor supplied time-mated animals are generally excellent. Therefore, the need to add excess animals per group to ensure the recommended number of pregnancies per group is less under current animal breeding practices from animal vendors. Purchasing vendor supplied time-mated animals also allows laboratories to conserve space and resources which would otherwise be required to maintain breeding colonies.

4.3.4 Maternal In-life and Caesarian Section Observations

The dosing period based on Gestation Day (GD) 0 by convention is defined as the day sperm or vaginal plug are observed in the rodent and for the rabbit the day observation of mating is confirmed. Based on day 0 of gestation, dosing for mice occurs GDs 6–15, rats 6–17 and GD 7–20 for rabbits.

In-life observations should include body weights (daily during the dose period), food consumption, and clinical observations (at least daily but more often as required).

4.3.5 Caesarian Section Examinations

Females are euthanized and examined before expected parturition GD 15, 18, 20/21 and 29 for hamsters, mice, rats and rabbits respectively. It should be noted rodents typically do not deliver early, therefore, if deliveries begin prior to scheduled examination this may be an indication of toxicity. Rabbits can abort with some frequency; however, this may be an indication of maternal toxicity which should be recorded. At laparotomy, ovaries are examined and the number of corpora lutea recorded. The uterus is removed and the intact uterus can be weighed but is only required for FDA studies if a food additive is tested. The uterine weight can be used to determine maternal body weight gain with and without the contribution of the uterine contents. Examination of the uterine contents includes the number of implantation sites, and the determination of the live, dead, and resorptions (early and late). Examination of the placenta should also be conducted and changes from normal noted. Uteri which appear nonpregnant should be examined either by pressing between glass slides or stained for implantations with 10 % ammonium sulfide to confirm pregnancy status (in some cases implantation occurs however implantations that fail to develop past the early embryonic stages begin to resorb and are not readily visible unless these examinations are conducted).

4.3.6 Fetal Evaluations

Live and dead fetuses should be identified with a tag after and placed in a compartmentalized container. All implantation sites should be recorded as to their status, i.e., live or dead fetus or early or late resorptions. Fetuses should be weighed and the sex determined along with the gross anatomical exam. Malformations and alterations are recorded and gross observations should be confirmed with subsequent examination in the visceral or skeletal evaluations. Table 1 reflects the study designs outlined in the preceding paragraphs.

Flexibility in the embryo–fetal study designs for the rodent studies allow for alternative paths for the soft tissue and skeletal exams. As per ICH guidelines, 50 % of rat fetuses from each litter are to be examined for visceral alterations. To perform craniofacial examinations, the heads from every other rat fetuses (approximately 50 %) in each litter should be removed and placed in fixative for later examination. Some laboratories conducting the rodent studies process one-half of the fetuses for the fresh visceral tissue (Staples Exam) or fixed tissue (Wilson's Exam) with the remaining half of the fetuses processed for skeletal evaluations. An alternative procedure is to evaluate all the fetuses with the fresh visceral examination and one half of the heads will be removed for Wilson's sectioning.

Table 1
Study design of definitive developmental toxicity (EFD) study

	Mouse	Rat	Rabbit
Strain or stock	CD-1	Sprague–Dawley or Wistar	New Zealand White, or Dutch Belted
Age on GD 0	10–14 weeks	10–14 weeks	4–6 months
Animals per group	22	20	20
Number of groups	4	4	4
Duration of dosing	GD 6–15	GD 6–17	GD 7–20
Physical observations	Twice daily	Twice daily	Twice daily
Body weights	Every other day after first dose	Every other day after first dose	Every other day after first dose
Food consumption	Weekly	Weekly	Daily during dose period
Caesarean section	GD 18	GD 20 or 21	GD 28 or 29
Fetal examinations (External/Visceral/Skeletal)	100/50/50 %	100/50/50 %	100/100/100 %

Adapted from Wise et al. [43]

All fetuses will then be processed for skeletal examination which requires staining and clearing to evaluate ossified bone and/or cartilage.

Fresh Visceral Examination (Staple's Technique)

Fetal fresh visceral dissections examine all the internal organs of the fetus and are well described by both Staples and Stuckhardt. This method provides an immediate evaluation of fetal outcomes. If sufficient resources and trained staff are available a quick evaluation of the study endpoints can be available in real time. There are several advantages using the fresh fetal dissection technique for the detection of fetal visceral alterations over the fixed tissue (Wilson's Method). Some of the advantages include reliability, thoroughness, accuracy, simplicity, and speed. Artifacts from autolysis, fixing, and slicing is reduced. The examination can be done immediately following caesarean section with minimum equipment and yields an intact skeleton which can subsequently be processed for skeletal examination. The fresh specimen and the natural coloration of in situ organs makes color photography of visceral alterations clear and concise. In addition, any lesion can be appropriately fixed for histopathological examination if necessary [44, 45]. Once the visceral exams are completed the fetuses are prepared for skeletal staining [46].

Wilson's Soft Tissue Sectioning Technique for Visceral Exam

Fetuses examined by the Wilson's soft tissue sectioning technique are first fixed in Bouin's solution, which is a mixture of saturated picric acid, formaldehyde, and glacial acetic acid. The purpose of the fixative is to fix the tissues, harden the soft tissues, and soften the bones in order to preserve the specimens and make it possible to slice them cleanly with a razor blade into sections of approximately 1 mm (or less) thickness. One drawback in this fixed tissue technique is that the original coloration of the tissues is lost and all tissues appear as a shade of yellow with the exception of blood which appears brown and the liver which appears olive green. The fetuses remain in the fixative for a period of 2 weeks and are then rinsed in alcohol prior to slicing. Since the formaldehyde and acetic acid fumes from the fixative are irritating and present a carcinogenic danger it is recommended that the rinsing and slicing be done under a fume hood or in a well ventilated area. As Bouin's solution stains tissues those working with these specimens should also wear latex gloves, preferably two pair.

The identity of the fetus is first verified after which it is given an external examination. The forelimbs are removed with either the razor blade or a sharp pair of scissors, and the head is sliced with between four and eight slices, the minimum being a slice at the beginning of the nasal passage just behind the nares, a slice at a point just before the eye bulges to examine the nasal passages (turbinates), the palate, and the nasal septum, a slice through the eye bulges to allow the examination of the lenses and upon removal of the lenses the retinas, and a slice behind the eye bulges at the level of the frontal/parietal suture in order to examine the lateral and third ventricles of the brain. If only these four slices are taken it is recommended that the brain be carefully removed from the last section to allow an examination of the cerebellum and meninges. If eight slices are taken, then the final slice should allow examination of the fourth ventricle of the brain. Following these slices the bottom of the head is removed from the fetus by slicing through the neck and the structures of the inner ear are examined.

The remainder of the fetus is sliced as thinly and evenly as possible and the sections laid on the wax block with the cranial side upward. The trachea, esophagus, and blood vessels are followed through the slices and the spinal cord and organs of the neck are examined. The right carotid, the right subclavian, left carotid, and left subclavian should be apparent and the right carotid and subclavian should be seen to converge to form the Innominate. Care should be taken to have a cut transect the top of the aortic arch. From the top of the aortic arch through the apex of the heart the technician should attempt to slice thinly enough to produce ten slices in which the following structures should be apparent: the aortic arch, the ductus arteriosus, the pulmonary artery, the atria, the semilunar valves, the tricuspid valve, the mitral valve, the ventricles, and the interventricular septum [47]. This fetal examination

method is less used by laboratories for several reasons. As noted, the fixative and processing time create laboratory hazards and delays in fetal examinations.

4.3.7 Skeletal Staining and Examination

Fetuses should be processed for skeletal staining. The procedures for staining and skeletal examination are explained and reviewed in several publications that provide excellent information for processing and examination procedures. The following citations provide excellent descriptions of staining procedures, evaluation and assessment recommendations and skeletal atlases also provide informative references for normal and abnormal skeletal anatomy [48–56].

The axial skeleton is examined after staining including the head (skull and facial bones), vertebral column (centra and arches), and rib cage (sternebrae and ribs). Skeletal alterations to this region include changes sometimes categorized by developmental toxicologists as malformations, variations, and delays in ossification, although categorization of structural changes as malformations versus variations is viewed as problematic by others.

A useful reference for the embryo–fetal development studies was published by Makris et al. in an updated Version of the Terminology of Developmental Abnormalities in Common Laboratory Mammals [57]. This excellent publication incorporates improvements and enhancements to both content and organization of the terminology to enable greater flexibility in its application, while maintaining a consistent approach to the description of findings. The revisions are the result of an international collaboration among interested organizations, advised by individual experts and the outcomes of several workshops. The terminology is organized into tables under the broad categories of external, visceral, and skeletal observations, following the manner in which data are typically collected and recorded in developmental toxicity studies. Only the commonly used laboratory mammals (i.e., rats, mice, rabbits) are addressed in the current terminology tables. The inclusion of other species that are used in developmental toxicity testing, such as primates, is considered outside the scope of the present update. Similarly, categorization of findings as, for example, "malformation" or "variation" remains unaddressed, in accordance with the overall principle that the focus of this document is descriptive terminology and not diagnosis or interpretation. The skeletal terms have been augmented to accommodate cartilage findings [57].

Regardless of the techniques used in fetal examinations, it is critical that personnel conducting the evaluations are experienced in the procedures and techniques used. The technical expertise required for the embryo–fetal development studies consists of an extensive knowledge of physiology and anatomy as well as a complete knowledge of fetal normal and abnormal anatomy. Therefore it is important for those conducting these studies be experienced and fully trained in the techniques and processes used to exam the fetuses.

4.3.8 Timing Considerations for Embryo–Fetal Developmental Toxicology Studies

The timing of DART studies relative to clinical trials and registration of pharmaceuticals is outlined in ICH M3 [58]. In spite of an effort to harmonize the preclinical studies to support clinical development for pharmaceuticals, regional differences remained in the expected timing of developmental toxicity studies to support the inclusion of Women of Child Bearing Potential (WOCBP) in clinical trials. In Japan, assessment of potential effects on female fertility and embryo/fetal development is required prior to the inclusion of WOCBP in any type of clinical trial. In the European Union, an assessment of potential effects on embryo/fetal development was expected prior to Phase I trials in WOCBP and female fertility studies were expected prior to Phase III. In the USA, WOCBP could be included in early, carefully monitored studies without reproductive toxicity studies, provided appropriate precautions were taken to prevent pregnancy. The recommended precautions included pregnancy testing, use of effective birth control, and enrollment after a confirmed menstrual period, with continued monitoring throughout the duration of the study. To further support this approach, informed consent would also include any known pertinent information related to reproductive toxicity, such as a general assessment of potential toxicity of pharmaceuticals with related structures or pharmacological effects. If no relevant information is available, the informed consent would clearly note the potential for risk. Over the past year, discussions to revise the ICH M3 guidance document have focused on efforts to better harmonize the preclinical requirements across all three regions. The current draft version of the ICH M3 guidance document (ICH M3(R2) July 15, 2008), states that the guidance provides general insight for biotechnology-derived products only with regard to the timing of preclinical studies relative to clinical development [41].

5 Prenatal and Postnatal Developmental Toxicity Studies

Assessment of potential effects on Prenatal and Postnatal Development (PPN) including Maternal Function is outlined in the ICH S5 (R2) document. These studies assess the hazard of both prenatal and postnatal exposure and are predominantly conducted in rodents (rat and mouse). In certain circumstances an alternate species may be necessary such as with thalidomide [59]. The PPN study is usually conducted very late in the clinical development program when most pharmaceutical companies expect to submit their drug for marketing approval. The Prenatal and Postnatal Development (PPN) study including Maternal Function examines Stages E and F as described in ICH S5(R2) and is the only study that dosing covers the periods of parturition through weaning [60].

5.1 Study Design Requirements

The designs of the PPN studies (and the drfPPN which is rarely done) are similar except for the number of mated females per group and the study durations. In most cases, information from the EFD studies is available prior to the start of a PPN study and in most cases; doses which are appropriate for the EFD study would also be applicable for the PPN study. Although information regarding the maternal and fetal response can be appropriate for selecting the dose levels, the continuation of the dosing throughout the parturition process and subsequent consequences on the survival and viability of the neonates is critical for a successful PPN study. However, assuming a drfPPN study was not conducted and the rodent EFD study demonstrated adequate maternal toxicity and no significant fetal toxicity, then appropriate dose levels could be selected in most cases. For other situations where a drfPPN study is deemed necessary, the selection of dose levels can be based on the prior information from developmental toxicity studies. Due to the uncertainty in the response in the early postnatal period, some laboratories may elect to utilize more than three drug-treated groups [60]. There is no ICH guidance for the design of drfPPN studies; however, a minimum but adequate range-finding study would include at least six mated females per group and at least three drug-treated groups.

The PPN study focuses on ICH S5(R2) Stages E and F to assess enhanced toxicity relative to that in nonpregnant females; prenatal and postnatal death of offspring; altered growth and development of newborns through weaning; and functional deficits in offspring, including behavior, maturation (puberty), and reproduction (F1 generation) [60].

5.1.1 Species Considerations

Rats or mice are typically used for the PPN study; however, as noted previously in some circumstances a more relevant species may be required [59, 61]. While the specific guidance for the PPN study states that the preferred species is the rat, the species must still be relevant to humans. The species selected is usually the species utilized for general toxicity studies. In general, outbred stocks of animals are used. While many laboratories obtain naturally mated females from vendors, which can be a more cost-effective alternative than maintaining a colony of male breeders to provide in-house mated females, this does have some disadvantages. The most obvious of these is the lack of any real acclimatization period prior to the initiation of dosing. Purchased timed-mated animals would generally be delivered on GD 1 or 2 with dosing then commencing on GD 6.

As with all the studies covered in this chapter, the most pharmacologically relevant species should be utilized for the PPN study [62, 63]. Use of alternative species other than the rat or mouse must be justified scientifically and special considerations for species specific differences in parturition and maternal care must be taken

into account. For example, the rabbit was used for the PPN study because of the sensitivity of this species to thalidomide, however special considerations for maternal cannibalization of newborns with defects had to be taken into account when designing the study [59].

The specifications outlined in the paragraphs below center on the use of the rat since this is by far the most commonly used species for the PPN study. If other species are required, the study design should reflect the physiologic and reproductive specifics of the species utilized for the study. As with all toxicology studies the experience of the laboratory conducting the study with the selected species should be a primary concern for study placement and conduct.

5.1.2 Number of Animals

ICH recommends rats for the PPN study, (note: the most pharmacologically relevant species should be used). Typically 20–25 females per group and three treatment groups and a control group are common. ICH recommends 16–20 litters per group and many laboratories choose to mate and assign at least a minimum of 20 dams per group or more to provide sufficient litters in each group. ICH S5(R2) guidance states that evaluation of between 16 and 20 litters per group for rodents tends to provide a degree of consistency between studies, and above 20–24 litters per group the consistency and precision are not greatly enhanced. These group sizes generally assure having these minimums for evaluation of the F1 generation during the growth and maturation phase and for evaluation at the time of F1 caesarean section, assuming appropriate dose selection [22].

These dams are identified as the F0 generation. This study design requires naïve males and females for mating and the females should be nulliparous. The strain of rat (or mouse) should be the same as the strain used for the other toxicology studies [22].

5.1.3 Dose Selection Considerations

The choice of the high dose should be based on data from all available studies (the EFD study, pharmacology, repeat dose toxicity and kinetic studies). The most common design for the PPN study is one vehicle control group and three dose levels for the test article groups. ICH S5(R2) states that some minimal toxicity is expected to be induced in the high-dose dams. The types of toxicity that may limit the high-dose level include effects such as increased or decreased body weight gain, target organ toxicity, exaggerated pharmacological response, or marked pup mortality in a preliminary study. Dose levels for use on a PPN study should not induce maternal deaths or body weight loss that extends for more than 2–3 days. ICH S5(R2) also establishes 1 g/kg/day as the limit (i.e., highest) dose for developmental and reproductive toxicity studies under most circumstances. If it is known or anticipated that 1 g/kg/day

will not produce minimal toxicity in the pregnant female, then the study design can include only one additional lower dose level of the test agent (i.e., three total groups). If the toxicity profile is largely unknown in a given species, then inclusion of four or more drug-treated groups may be warranted. Also if the toxicology profile is unclear (which should be rare at this point in development) a dose range-finding (drf) PPN study should be run prior to the definitive study.

In most cases doses from the EFD study provide a scientific basis for selection of doses for the PPN study since the dose period for the EFD study covers GD 6–17 (implantation through closure of the hard palate) cover the majority of the dose period during gestation for the PPN study (PPN study begins dosing GD6 and continues through parturition typically GD 21–22). Observations of malformations in the EFD study should be considered when designing the PPN study. Examples of life threatening malformations such as interventricular septal defects, and exencephaly observed in the EFD study will impact the litter observations in the PPN study. Fetuses with serious malformations once delivered will not survive following parturition and most likely will be cannibalized by the mother postpartum. If a dose results in sufficient reductions in litter size or whole litter loss it is possible the number of surviving pups to adequately assess survival and growth and development of the neonates will not be available.

For some drugs using the data from the EFD study or the repeat dose toxicology studies cannot adequately predict effects on parturition, pup survival, or maternal care of the litter postpartum. Other pharmacologic classes of drugs are known to have potential effects on parturition, maternal litter care, lactation, or pup survival and growth. An example of a drug that exhibits toxicity selectively after delivery is the angiotensin II antagonist, Losartan, which had no effects on fetal development but resulted in lower pup weights, delays in development, and preweaning and post-weaning deaths in the PPN study [64, 65]. In addition, anti-inflammatory drugs, administered orally to pregnant rats during the 19th to 21st day of gestation can affect the time of onset of parturition (TOP), duration of parturition (DOP), bleeding during parturition, and perinatal mortality. Acetylsalicylic acid, salicylic acid, and cortisone acetate (in high dosages) all delay significantly the (TOP) when compared to controls. The acid non-steroids prolong the parturition time, and a marked increase of bleeding at parturition was evident in the acid nonsteroid treated group with only a slight increase in cortisone treated animals. Acetylsalicylic acid, salicylic acid, and cortisone did not affect the number of pups born dead/total pups, but sodium salicylate and acetylsalicylic acid resulted in an increase in the number of fetal deaths [66].

Recognizing that even with a considerable knowledge base of the toxicity of a given compound, it may be difficult to assess a drug's toxicity to late term gestation or neonatal survival and development. Because much of the maternal weight gain in the late phase of gestation (GD's 18–21) is actually fetal weight gain; it may be advisable to maintain the dose based on the body weight of the dam on GD 17. As most compounds are administered by volume based body weight with a drug dosage set at mg/kg consideration should be given to maintaining the dose administered based on the body weights obtained from gestation days 15–17 to minimize overdosing the dams.

With the examples presented above it is apparent potential effects on late stage gestation, parturition, maternal care for the litter, milk production and neonatal growth and development can be adversely affected by drug exposure. If the pharmacologic class of the drug under consideration has known effects in late gestation or lactation, a drfPPN study is advisable before the definitive study is undertaken. While the drfPPN study will add some amount of time to the development program, conduct of this study should provide an insight to the potential toxicities which may be observed in the definitive study. Given the ICH requirements for minimum numbers of litters and therefore pups in the F1 generation, it is advisable to not chance the loss of whole litters or compromised pups due to dosages that are excessively toxic to the dam or neonate.

ICH S5(R2) states that if a "no observed adverse effect level" for "reproductive aspects" is not identified in a PPN study, then additional studies may be needed. The designs of those additional studies will be compound specific and should be designed with the pharmacologic class of the test article in mind. Ideally, either the low- or mid-dose level should demonstrate no adverse developmental toxicity. While ICH S5(R2) defines the various "adverse effects to be assessed" (e.g., prenatal and postnatal death of offspring, altered growth and development), the guidance document did not attempt a definition for the adjective "adverse." Whether or not a developmental effect is adverse depends on a number of parameter-specific factors, which can best be addressed by the scientists conducting the study and/or those persons most knowledgeable about other aspects of the compound.

The spacing between dose levels is also an important decision. **Note 8** states that dose-responses may be steep, and wide intervals between doses would be inadvisable. Narrow intervals may also pose problems when the variability of a toxicological response overlaps between adjacent dose levels. A minimum multiple of twofold and a maximum of four- to fivefold between dose levels are usually adequate.

5.1.4 Exposure Timing Considerations and Treatment Period

Females are mated and assigned to one of four groups as described previously. It is advisable not to place all females into mating at the same time unless sufficient resources are available to dose, observe and manage all the littering observations. The reason for this is to space out, if possible, the number of GD0 dams so that a large percentage of the number of animals are not littering on the same day. During the lactation period there are numerous and very labor intensive observations and tests and staggering the number of dams beginning dosing and eventually littering will allow adequate time for all of the study procedures which are required. The F0 females are exposed to the test substance from implantation to the end of lactation (i.e., stages C–E). The accuracy of the time of mating should be specified since this will affect the variability of fetal and neonatal parameters. Similarly, for reared litters, the day offspring are born will be considered as postnatal or lactation day 0 (LD_0) unless otherwise specified. However, particularly with regard to delays in, or prolongation of parturition, reference to a postcoital time frame may be useful. Dosing of the females begins on Gestation Day (GD) 6 and continues throughout parturition and through Lactation Day (LD) 20. Once daily dosing is the most common regimen; however, the toxicokinetic characteristics of the drug should be taken into consideration. In some cases compounds with short half-lives twice daily dosing may be considered. For compounds with long half-lives a modified dosing schedule may be appropriate. In either case, scientific justification for either alteration in the dosing schedule should be used in planning the dosing regimen. Employing a drfPPN study in these cases can be a useful tool to assess if effects on parturition or pup survival may be expected in the definitive study [22, 43, 62].

As noted previously, maternal weight gain in the late phase of gestation (GD's 18–21) is predominately due to fetal weight gain. Therefore adjusting the dose volume late in gestation to match the body weight changes of the dam may not be advisable. Maintaining the dose based on the body weight of the dam on GD15–17 where data exists on maternal toxicity from the EFD may be advisable for most test articles. Since the dose volume is adjusted based body weight change keeping the same dosage on a mg/kg dose as administered on GD15–17 should minimize overdosing the dams in late gestation [22, 43].

5.1.5 Maternal In-life Observations F0 Generation

Animals should be visually inspected for moribundity and mortality twice daily following treatment initiation during gestation and lactation. Post-dosing observation should be done around the expected T_{max}. Body weights should be recorded at confirmation of mating and at randomization into dose groups, then on GDs, 4, 6, 12, 15, 18, and 20 and on LD 0, 4, 7, 14, and 21 at a minimum. ICH S5(R2) suggests that weighing the females twice weekly is acceptable but more frequent recording of body weight provides

for accurate adjustment of the dose volume. Food consumption should be measured to correspond with the times of body weight measurement prior to and after initiation of treatment as noted for body weights. Pregnant females should have food consumption measured during gestation and lactation. Measurement of food consumption later in the lactation period does not necessarily reflect the maternal consumption as the pups are beginning to consume solid food (usually by LD 13–14). Food consumption is a valuable measurement even late in the lactation period as decreased food consumption may indicate maternal and/or pup effects due to the test article, however, decreases in consumption should be evaluated in the context of both F0 and F1 body weight changes.

Observations of the dams should be increased as the anticipated parturition date nears. It is recommended to begin more frequent inspections for evidence of littering beginning on GD 18. F0 females should be examined 2× daily for delivery and possible dystocia. Nesting behavior is usually exhibited as the dam's parturition nears. While not specified in ICH S5(R2), observation of nesting behavior or an absence of such can indicate maternal toxicity and dams displaying an absence of nesting behavior should be observed closely for evidence of dystocia. Similarly, maternal care and nursing behavior are important observations to evaluate possible maternal toxicity. Evidence of maternal cleaning and nesting behavior including nursing of pups is the norm; however, the absence thereof indicates maternal toxicity and may result in the loss of neonates if maternal cleaning and care does not occur quickly following parturition. Maternal care and nursing behavior is a critical assessment throughout the lactation period. Maternal care is most critical during the first few days following delivery. Neonatal losses occur predominately during the first 4 days of lactation. Observations during the lactation period should include non-maintenance of the nest and little to no milk in the pups' stomach visible through the skin. These observations may aid in assessing if pup mortalities are maternally mediated due to poor maternal care rather than a direct effect upon the offspring.

Toxicokinetic assessments will have been conducted as part of the EFD study so if the PPN dose levels are similar, then those data would cover the pregnant phase of the PPN study. Despite the changes in the maternal condition relating to the milk production and hormone changes during the lactation period, it is uncommon to collect further samples during this period [43].

At weaning F0 dams are necropsied and if target tissues have been identified from previously conducted toxicology studies these tissues may be included for histopathologic examination. A gross (macroscopic) examination is conducted for the thoracic and abdominal viscera. Organs with macroscopic findings and samples

from concurrent control females are preserved in an appropriate fixative for possible histological evaluation. Mated females that do not deliver by presumed GD 24 should be euthanized and examined (uteri should be briefly immersed in ~10 % ammonium sulfide solution in order to visualize early implantation sites). In females that have littered, the number of implantation sites (as represented by metrial glands) should be recorded to enable a comparison with the number of pups born and allow an assessment of the in utero and neonatal loss (when females deliver at night, dead or malformed pups are often cannibalized by the dam before the first inspection of the litter) [60].

5.1.6 Parturition and Preweaning Litter Observations F1 Generation

Observations of parturition including the onset and completion of delivery and any signs of difficulty in parturition should be noted when observed. Based on previous experience the majority of Sprague–Dawley females will initiate parturition during the lights-on period with delivery completion occurring rather quickly (approximately 2 h). For determination of the transition from gestation to lactation it is acceptable to assign a whole day value (e.g., GD 21.0 or 22.0) to females that have completed delivery at the first observation of the day. Females that complete delivery during the workday are assigned a half-day value (e.g., 21.5 or 22.5) [43]. The day of parturition is considered day 0 of lactation (LD_0). The number of live and dead pups born in each litter is recorded following completion of parturition along with external abnormalities and the sex of each pup should be recorded. Pups should be counted daily until weaning and they should be given a detailed examination on the day body weights are taken. Individual pup body weights are usually recorded on lactation days 1, 4, 7, 14, and 21. For studies where there is a high incidence of pup deaths or excess toxicity during lactation, more frequent observations and weighing may be necessary.

In the event excess maternal toxicity or neonatal toxicity or deaths are encountered in a PPN study, such as whole litter loss, lack of maternal care resulting in excessive pup deaths, or marked decreased pup body weight gains in the high-dose group, a fostering/cross-fostering study may be necessary. An additional study can be conducted to clarify if the effects are due to exposure during gestation and/or lactation (i.e., for better assessment of human relevance). This study design incorporates the same parameters as the standard PPN study but only two groups (vehicle control and high-dose) are used in this design. Dams are treated for the same duration as in the PPN study. On the day of birth, all pups are transferred to different mothers in the following combinations: control pups to both control and high-dose dams, and high-dose pups to both control and high-dose dams. The combination of control pups cross-fostered to the high-dose dams examines the

effects of lactation-only exposure, while the combination of high-dose pups to control dams examines the effects of gestation-only exposure.

While not a requirement, many laboratories cull the litter size usually on LD4 to 4 or 5 pups per sex to standardize litter size and decrease variability based on large differences in litter size. For studies where maternal toxicity or neonatal loss reduces the litter size culling may not be needed or desired. On LD4 if the procedure is to cull the litters the selection of pups should be random for both pups removed and those selected to remain on study.

Current consensus is the best indicator of physical development is body weight and **Note 21** of ICH S5(R2) states that the acquisition of preweaning developmental milestones is highly correlated with the body weight and that this weight should be related to postcoital time rather than postnatal time, especially where significant differences in gestation length occur. This does not preclude the monitoring of physical milestones and reflex acquisition but due to this close correlation many laboratories utilize body weight alone as the indicator of postnatal growth and development. Physical parameters often utilized include pinna unfolding, hair growth, incisor eruption, and eye opening. When reflex ontogeny is included, it is usual to include at least two measures of reflex ontogeny and these may include surface righting, auditory startle, negative geotaxis, air righting, Preyer response, or pupillary reflex. The option also exists to monitor more complex forms of behavior and a good example of this is the ontogeny of swimming behavior, for which seven different stages can be clearly distinguished and easily evaluated in rats and mice [67]. Preweaning developmental landmarks which have routinely been evaluated in this study such as: righting reflex, pinnae detachment, and eye opening are more correlated with body weight gain than they are developmental age. These parameters can be evaluated; however, the timing the milestone is achieved is variable and some laboratories opt to not assess these landmarks in the PPN study [58, 68, 69].

5.1.7 Postweaning Assessments

At weaning (around PND 21) pups from all available litters are selected randomly one per sex per litter to provide 20 males and females per dose and control group for further examinations with the exception that pups with obvious abnormalities or delayed growth (i.e., animals with a body weight more than two standard deviations below the mean pup weight of the respective litter) should not be included, as they are unlikely to be representative of the treatment group.

Two developmental landmarks should be evaluated routinely postweaning. The study should assess the onset of puberty in both sexes and **Note 21** of ICH S5(R2) strongly suggests monitoring vaginal opening of females and cleavage of the balanopreputial

gland (i.e., preputial separation) of males for this assessment [68, 70]. Testis descent is not a recommended method of assessment as this can be variable and very subjective in comparison to balanopreputial separation. The latter is associated with increasing testosterone levels whereas testis descent is not [68]. As with other physical milestones, these parameters are affected by general growth and therefore it is recommended that the body weight of the animal be recorded at the time of attainment to differentiate between specific compound effects and those related to growth. Some laboratories examine animals every other day for attainment of these landmarks. Vaginal opening is evaluated for the F1 females beginning from Day 30 until complete and preputial separation for the F1 males from Day 40 until complete [68].

These pups are identified as the F1 generation and the day the pups are weaned and selected to continue on study is designated as postnatal day 21 (PND_{21}). On PND 21, the selected F1 pups are randomly assigned to one of three cohorts of animals, as follows:

- Reproductive/developmental toxicity testing (Cohort 1): at approximately 12 weeks F1 males and females are selected for mating (nonlitter mates). Evidence of mating is obtained as described previously. Observations of reproductive outcome can be done by examination of mated F1 females on presumed GD13 or the females can be allowed to deliver naturally with litter assessments conducted as described for the F0 generation.
- Developmental neurotoxicity testing (Cohort 2): Behavior and Learning assessments described in following section [71].

It is suggested in **Note 19** of ICH S5 (R2) that at the time of weaning (i.e., PND 21), one animal per sex is selected from each litter. While it is feasible to conduct all subsequent evaluations using only these animals, the option exists to randomly select more than one animal/sex/litter. Doing so allows the option of having naïve animals for at least one behavioral test and they could also provide animals for a second mating trial for reproductive performance should issues be encountered during the first mating.

If multiple assessments are planned as described above it is recommended for separate cohorts to be selected for the F1 pups as outlined:

- *Reproductive/developmental toxicity testing* (*Cohort 1*):
  - Cohort 1A: One male and one female/litter/group (20/sex/group): priority selection for primary assessment of effects upon reproductive systems and of general toxicity.
  - Cohort 1B: One male and one female/litter/group (20/sex/group): priority selection for follow-up assessment of reproductive performance by mating F1 animals, and for obtaining additional histopathology data in cases of

suspected reproductive or endocrine toxicants, or when results from cohort 1A are equivocal.

- *Neurobehavioral Evaluations/Behavior and Learning* (*Cohort 2*):
  - Cohort 2A: Total of 20 pups per group (ten males and ten females per group; one male or one female per litter) assigned for neurobehavioral testing followed by neurohistopathology assessment as adults.
  - Cohort 2B (optional): Total of 20 pups per group (ten males and ten females per group; one male or one female per litter) assigned for neurohistopathology assessment at weaning (PND 21 or PND 22). If there are insufficient numbers of animals, preference should be given to assign animals to Cohort 2A.

5.1.8 Behavior and Learning Assessments

ICH S5(R2) recommends evaluation of functional tests of F1 behavior. ICH does not define specific tests to allow investigators to choose testing methods best suited to evaluate sensory function, motor activity and learning and memory. The best evaluation methods and the timing of the tests conducted to assess sensory function, motor activity and learning and memory must be selected based on good science, the species being tested and the experience of the laboratory conducting the evaluations. Evaluations typically are conducted after weaning and before the mating of the F1 generation. For many of the tests, commercially available equipment and procedures are available and many laboratories have databases whereby historical baselines can be used to assess current studies. Utilization of automated data collection equipment has a further advantage in that observer bias and variability in observations are minimized by use of electronic data collection automated equipment using standardized procedures and validated equipment. A good source of information on the various appropriate tests available is contained within the EPA Developmental Neurotoxicity Guideline (OPPTS 870.6300) and the OECD Developmental Neurotoxicity Guideline (OECD 426), the investigator should be confident of the validation of the test within the laboratory and the tests should be capable of detecting both increases and decreases in ability [63, 72, 73].

Motor activity is evaluated with equipment which utilizes a chamber with photobeam detectors which electronically record animal movements by recording when the photobeams are broken using an automated activity monitor. While there is no standard timeframe required by regulators, testing around PND 42 is common. These evaluations are designed to evaluate initial response to the novel environment of the activity monitor, as well as to observe the rate of habituation to that environment over the test periods. Sensory function is commonly evaluated by the acoustic startle response. This test is based on an automated system also where

the test subject is placed in an acoustic chamber on a platform system to measure animal response to an auditory stimulus. The test evaluates the startle response to an auditory stimulus as well as prepulse inhibition and habituation to an acoustic startle stimulus typically assessed on or around PND 43. This test typically is conducted using multiple trials (a set of 70 trials is common) using multiple chambers capable of digitally recording the startle responses as well as habituation of the response [63, 72, 73].

Evaluation of learning and memory typically is conducted using one of two test paradigms. The first type of test evaluates learning and memory by passive avoidance where the animal learns to enter the compartment to avoid receiving a mild but unpleasant electric shock. A second commonly used learning and memory assessment is a maze test. The swimming water maze is often used due to the advantage this test has in evaluating both motor and learning and memory development in the same test.

As noted, the tests selected for assessment of motor and sensory function and learning and memory. Basic principles should be taken into account regardless of the test selected.

- Quantitative assessments are preferred to eliminate inter and intra observer bias and variability;
- The laboratory conducting the test should have documented experience and operational procedures along with validated test methods. Test systems should always be evaluated for capability to identify known toxicants for the parameters being evaluated;
- Variability is acceptable, however, controls of variability should be in place and the laboratory should have adequate historical control databases to allow scientifically sound evaluation of the data for normal variability within the study being evaluated.
- In cases where subjective evaluations are necessary, variability between observers and variability in singular observers over multiple timepoint evaluations should be controlled as much as possible. Laboratories should demonstrate different observers are capable of evaluating behaviors consistently with minimal variability [63, 72, 73].

5.1.9 Assessment of F1 Fertility and Reproductive Performance

ICH S5(R2) guidance does not specify or detail how the fertility of the F1 generation should be assessed, however, it is generally understood that an assessment based on the guidance for the fertility study should be followed. In some cases, however, it may be more appropriate to allow the F1 generation to litter and assess the pregnancy and lactation performance of the F1 together with the survival growth and development of the F2 generation up to LD 7. This may be particularly appropriate where significant problems have been observed in the pregnancy performance of the F0 generation or the sexual development and fertility of the F1

generation females [43]. Although not required or suggested by ICHS5 (R2), the monitoring of estrous cycles in the F1 generation females prior to pairing may be considered if the test article is suspected of influencing the reproductive system or if triggered by precocious or delayed onset of puberty. This would normally be conducted by assessment of vaginal cytology by means of vaginal smears for a period of between 10 and 14 days immediately before pairing commenced. Reproductive performance assessment should be conducted when the F1 generation has reached sexual maturity (approximately 70 days of age or postnatal week 10). If poor fertility has been observed, which by re-mating with naïve females is considered a male-mediated effect, then histopathology and/or sperm evaluations may be performed at termination of the F1 males [43].

5.2 Bioanalytical and Toxicokinetic Considerations for Studies to Evaluate Fertility, Developmental Toxicity, and Prenatal and Postnatal Developmental Toxicity

ICH S3A addresses analysis of pharmacokinetic data on the repeat dose and reproductive and developmental toxicology studies. The repeat dose studies can provide valuable pharmacokinetic data for rodents since they are commonly used for both the repeat dose and developmental and reproductive toxicology studies. Pharmacokinetic information in pregnant or lactating animals may be required and in many cases dose-range finding studies are advisable to determine if pregnancy and lactation will affect the ADME parameters for the compound evaluated in the Fertility, Developmental, and Prenatal and Postnatal Developmental Toxicology studies [74].

The limitation of exposure in reproductive toxicity is usually governed by maternal toxicity. Toxicokinetic monitoring in reproductive toxicity studies may be valuable in some instances, especially with compounds with low toxicity, but these data are not needed for all compounds.

Where adequate systemic exposure might be questioned because of absence of pharmacological response or toxic effects, toxicokinetic principles could usefully be applied to determine the exposures achieved by dosing at different stages of the reproductive process. A satellite group of female animals may be used to collect the toxicokinetic data.

For the Fertility Study, the general principles for repeated-dose toxicity studies apply. The need to add bioanalytical and pharmacokinetic evaluations will depend on the dosing regimen used and the information already available from the repeat dose studies in the selected species [74, 75].

For the EFD and PPN studies, the treatment regimen during the exposure period should be selected on the basis of the toxicological findings and on pharmacokinetic and toxicokinetic data from the repeat dose and dose range-finding studies. Consideration should be given to the possibility that the kinetics will differ in pregnant and nonpregnant animals. Toxicokinetics may involve exposure assessment of dams, embryos, fetuses or neonates at

specified days in each study design. Secretion in milk may be assessed to define its role in the exposure of newborns. In some situations, additional studies may be necessary or appropriate in order to study embryo/fetal transfer and secretion in milk. Additional consideration should be given to the interpretation of reproductive toxicity tests in species in which placental transfer of the substance cannot be demonstrated [74].

Exposure is represented by pharmacokinetic parameters demonstrating the local and systemic burden on the test species with the test compound and/or its metabolites. The area under the matrix level concentration-time curve (AUC) and/or the measurement of matrix concentrations at the expected peak-concentration time C_{max}, or at some other selected time $C_{(time)}$, are the most commonly used parameters [74, 75].

Other measurements, for example urinary excretion, may be more appropriate for some compounds. Other derived parameters, for example bioavailability, half-life, fraction of unbound drug and volume of distribution may be of value in interpreting toxicokinetic data. Thus, the selection of parameters and time points has to be made on a case-by-case basis considering the general principles as outlined in Section 3 [74].

Increasing exposure may also occur during the course of a study for those compounds which have a particularly long plasma half-life. Careful attention should also be paid to compounds which achieve high C_{max} values over comparatively short time periods within the dosing interval. Conversely, unexpectedly low exposure may occur during a study as a result of auto-induction of metabolizing enzymes [74, 75].

The ICH S5(R2) and the ICH S3A Guidances do not require assessment of neonatal exposure, however, the latter guidance document states "Toxicokinetics may involve exposure assessment of dams, embryos, fetuses or newborn at specified days. Secretion in milk may be assessed to define its role in the exposure of newborns. In some situations, additional studies may be necessary or appropriate in order to study embryo/fetal transfer and secretion in milk." [74]

Interspecies differences in metabolism can lead to differences in developmental and reproductive toxicity. Understanding metabolic pathways in test species and humans is important to designing and interpreting animal studies and their potential relevance to humans. Interspecies differences in embryonic development and the extra-embryonic environment can also lead to species discordance in developmental toxicity. Toxic effects on the visceral yolk sac can lead to malformations in rodents that would likely be of reduced concern for humans. The importance of toxicokinetics in study design and interpretation should take into account species selection, dosing regimen, and extrapolation of data to humans (i.e., relative exposure, safety margins). For pharmaceutical evaluation if

possible the route of human exposure should be used if possible as the route of administration in animal studies. Alternative routes of exposure may be used (e.g., IV), as long as the dosing regimen provides an exposure profile representative of that expected in humans. Characterization of pharmacokinetics (PK) by various routes will allow design of a dosing regimen that will produce comparable exposures and provide appropriate coverage of gestation [40, 75].

5.3 Data Presentation for Prenatal and Postnatal (PPN) Studies

There are no specific requirements detailed in the ICH S5(R2) guidance document regarding the presentation of data; however, the guidance does recommend "tabulation of individual values in a clear concise manner to account for every animal that was entered into the study." [22] Data should be presented to allow the reader to follow individual animals from the maternal data and through birth, lactation, and through the preweaning and postweaning periods for the F1 generation. While the litter is the experimental unit for evaluation in the PPN study as with the fertility and EFD studies, all the tables of individual findings should identify the F0 female/litter or the selected F1 males and females in the PPN study since there are individual animals assessed in the F1 generation.

Graphical presentation of continuous data (i.e., values for data collected on multiple days or within a specific time frame such as mean maternal body weights, F1 generation body weights, activity scores across time, or startle response across multiple trials) is often more clearly visualized in a line graph. Summary tables of all major categories of data are essential but the individual data that contribute to these means must be in a form that makes the review of the data easy. Mean absolute body weights, body weight gains, and mean food consumptions are generally included. Summary and individual tables of physical observations may also be included. Summary tables for data collected at littering and during lactation would include parameters such as postimplantation survival, live pups delivered; live pups per litter on PNDs 0, 4, 7, 14, and 21, live birth index. For preweaning developmental milestones and reflex acquisition data should be calculated per litter, and not the individual pup, as the unit of measure [43].

Parametric analyses are performed for body weights, food consumption, litter size, gestation length, precoital intervals, and attainment of developmental landmarks. Nonparametric measures should be used for preimplantation loss, postimplantation loss, pup postnatal survival, and sex ratios. Behavioral data by nature variable due to inherent individual animal (or observer) variability should be analyzed by repeated measure analyses to increase the power of the test to detect a change [43, 60, 76–78].

6 Evaluating Biopharmaceuticals for Reproductive and Developmental Toxicity in Traditional Models

Biopharmaceuticals include recombinant proteins and peptides, and antibodies that are designed to be highly specific for their targets in the human body. While the principles of testing are similar for small molecule pharmaceuticals and large molecule biopharmaceuticals, the practices of preclinical safety assessment often differ in large part due to the inherent differences in product attributes between small molecules and large molecules [58]. The principles of DART testing for biopharmaceuticals are outlined in International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) document S5(R2). However, because many biopharmaceuticals are species-specific, alternate approaches may be needed to evaluate DART potential as outlined in ICH S6. For molecules that show species-specific cross-reactivity restricted to nonhuman primates (NHP), some aspects of DART may require NHP testing. For biopharmaceuticals that are uniquely specific and only active on intended human targets or human and chimpanzee targets, surrogate molecules that cross-react with the more traditional rodent species may need to be developed and used for DART testing. Because biopharmaceuticals are designed to be specific for their human target, they often show limited cross-reactivity with traditional toxicology species used for DART testing (rat and rabbit) and, therefore, nontraditional approaches need to be implemented for toxicity testing, including DART testing. Also, whereas small molecules (~<1000 Da) and their metabolites can diffuse across plasma membranes and the placenta by simple diffusion, large molecules (including biopharmaceuticals) do not freely diffuse across membranes, so their distribution is largely limited to the vascular space. The diffusion of molecules across membranes depends on molecular size, lipid solubility, degree of ionization, and plasma protein binding [79, 80]. Because small molecule pharmaceuticals have the potential to diffuse across plasma membranes, the gametes and embryo/fetus can be exposed to the pharmaceuticals at any time during reproduction and development [58]. In contrast, large molecule biopharmaceuticals do not appreciably diffuse across plasma membranes, including the placenta and, therefore, have limited access to the conceptus [80]. Due to limited diffusion, large molecule biopharmaceuticals are not expected to gain access to the intracellular environment and, therefore, are not likely to interact with DNA or other intracellular organelles. For these reasons, large molecule biopharmaceuticals need to be considered differently from small molecules when evaluating DART risks. However, certain types of large molecules can cross the placenta by specific active transport mechanisms. For example,

antibodies are transported across the placenta by Fc receptor–mediated endocytosis [81, 82]. Alternatively, genetically modified transgenic animals may also need to be considered. Surrogate molecules and transgenic animals may also be considered for DART testing even if the biopharmaceutical is active in NHPs in order to reduce the use of NHPs [58]. Because of the unique properties of biopharmaceuticals, a case-by-case approach is needed for DART and general toxicity evaluation, which requires consideration of specific product attributes including biochemical and biophysical characteristics, pharmacological activity, and intended clinical indication [58].

The ICH S6(R1) Guideline Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals indicates that biotechnology-derived pharmaceuticals often preclude standard toxicity testing designs in commonly used species. Therefore, safety evaluation programs should include the use of relevant species which is defined as the one in which the test material is pharmacologically active due to the expression of the receptor or an epitope (in the case of monoclonal antibodies) [58]. The need to conduct reproductive and developmental toxicity studies is dependent upon the biopharmaceutical itself, clinical indication and intended patient population. The study design and dosing schedule should reflect to the extent possible the intended clinical use and route and may be modified based on issues related to species specificity, immunogenicity, biological activity and/or a long elimination half-life. For example, concerns regarding potential developmental immunotoxicity, which may apply particularly to certain monoclonal antibodies with prolonged immunological effects. In some cases, immunogenicity or the pharmacokinetic parameters of the molecule may require special study designs modified to assess immune function of the neonate. If the rodent or rabbit is a relevant species, the timing of reproductive toxicity studies outlined in ICH M3(R2) Guideline should be followed for the timing of data on fertility for products where rodents are relevant species [83, 84].

Immunogenicity can be an issue when rats and rabbits are chosen for toxicology testing of human proteins. Strategies to overcome immunogenicity have been used successfully in DART studies and have been most successful in the rat and rabbit EFD studies since the treatment period is very short (implantation through organogenesis). The dose and treatment schedule can be adjusted by administering a higher dose and/or dosing more frequently to overcome the effects of anti-drug antibodies [58]. An example is the anti-TNFα receptor Fc fusion protein etanercept was shown to neutralize rat and rabbit TNFα and, therefore, the rat and the rabbit were considered pharmacologically relevant species for DART testing. Longer dose periods detected neutralizing antibodies in rats and rabbits after 2–4 weeks of dosing. The EFD study was conducted using a daily dosing regimen at high-dose levels to

minimize the impact of immunogenicity. However, fertility and prenatal and postnatal development studies, which require longer-duration dosing, were not conducted because the neutralizing antibody response would have reduced exposure, and thus pharmacology, to the biopharmaceutical over time [58].

If a protein is highly immunogenic, it may still be possible to assess biopharmaceutical-related effects from implantation through organogenesis by dosing animals during different windows of development (e.g., if the protein was immunogenic only after 10 days in the 2-week toxicology study, the group size could be doubled and half the animals could be dosed for 1 week beginning at implantation and the second half of the group dosed during the last week of organogenesis) [58]. This type of dosing paradigm is often used when trying to determine the sensitive period for a teratogen. It is also important to note that dividing the exposure into windows of development may change the dose–response or pattern of developmental effects [85, 86]. Shorter periods of exposure may produce less severe effects on embryo–fetal growth and lethality, while enhancing the ability to detect malformations. If embryo–fetal lethality is due to or occurs concurrent with dysmorphogenesis, reducing the exposure interval (duration) may improve embryo/fetal survival to manifest increased malformation rates [86].

A similar cohort study design could also be considered for fertility and pre/postnatal development studies for immunogenic proteins. Although immunogenicity is a particular concern for rodent and rabbit studies, it may also be a concern in higher-order species including NHPs, requiring that similar dosing strategies be applied [58, 86, 87].

6.1 Considerations of Placental Transfer in Traditional Species

Maternofetal transfer of immunoglobulins between species is highly variable with extensive gestational transfer of maternal immunoglobulins in primates (including humans) via the chorioallantoic placenta as well as in rabbits and guinea pigs via the inverted yolk sac splanchnopleure. Neonatal rodents (rats and mice) receive passive immunity predominantly postnatally. This transfer is mediated principally via FcRn receptors [58]. Therapeutic monoclonal antibodies (mAbs) of the IgG1 subclass are transported most efficiently to the fetus. In all animal species used for testing developmental toxicity, fetal exposure to IgG is very low during organogenesis, but increases during the latter half of gestation such that the neonate is born with an IgG1 concentration similar to the mother (but not rats and mice). Pregnancy outcome data from women gestationally exposed to mAb is limited. In general, the findings are consistent with the expected low exposure during organogenesis. Guinea pigs and rabbits are potential candidates as “alternatives” to the use of nonhuman primates as the maternofetal transfer in the last part of gestation is at a level similar in humans [88]. Based on the

pattern of placental transfer of IgG in humans, study designs that allow detection of both the indirect effects in early gestation plus the effects of direct fetal exposure in mid and late gestation are recommended for developmental toxicity of mAbs. It is important to note that placental transfer of antibodies (mAb and Fc-fusion proteins) differs between rodents and primates [88]. In humans (and in NHPs) transfer of antibodies across the placenta occurs primarily during the latter part of pregnancy, i.e., after organogenesis [87, 89]. This also appears to be the case for rabbits [80].

If the objective of a rodent development study is to mimic clinical exposure relative to fetal development, it may be necessary to adjust the dosing period to provide exposure during critical periods of development. For example, the rodent immune system is developmentally immature at the time of birth relative to the primate immune system [90]. Consequently, certain aspects of immune system development that occur postnatally in rodents occur prenatally in primates. In order to mimic effects of human in utero exposure during the fetal period (i.e., the second and third trimesters), it would be necessary to treat the pregnant rodents starting at the end of organogenesis and to continue dosing during the postnatal lactation period. In rodents, antibodies are effectively transported in the breast milk and are absorbed across the neonatal gut by FcRn. In humans and NHPs, IgG antibodies are transported predominantly across the placenta by FcRn with only IgA antibodies being transported to any great extent in the human breast milk [91]. With these differences in mind, strategies can be implemented that mimic human exposure even though the dosing intervals and routes of exposure may differ between the species.

Despite embryo/fetal exposure differences between rodents and primates, the rodent is a useful model in evaluating effects of mAbs on the maintenance of pregnancy and potential effects on the pregnant female. For non-antibody biopharmaceuticals, which are not transported across the placenta, species differences in the placenta are less important and, therefore, rodents are relevant species for developmental studies as long as they are pharmacologically relevant.

When no relevant animal species exists for testing the clinical candidate, the use of transgenic mice expressing the human target or homologous protein in a species expressing an ortholog of the human target can be considered, assuming that sufficient background knowledge exists for the model (e.g., historical background data). For products that are directed at a foreign target such as bacteria and viruses, in general no reproductive toxicity studies would be expected [83].

For the reproductive and developmental toxicology studies, the drug product that is used in the definitive studies should be comparable to the product proposed for the initial clinical studies. However, with biopharmaceuticals change is inherent in the

development program as the manufacturing processes are modified to improve product quality and yields. The potential for such changes to impact the extrapolation of the animal findings to humans must be considered. The comparability of the test material during a development program should be demonstrated when a new or modified manufacturing process or other significant changes in the product or formulation are made in an ongoing development program. Comparability can be evaluated on the basis of biochemical and biological characterization (identity, purity, stability, and potency). In some cases additional studies may be needed (i.e., pharmacokinetics, pharmacodynamics and/or safety). The scientific rationale for the approach taken should be provided [83].

When the weight of evidence (e.g., mechanism of action, phenotypic data from genetically modified animals, class effects) suggests that there will be an adverse effect on fertility or pregnancy outcome, these data can provide adequate information to communicate risk to reproduction, and under appropriate circumstances additional nonclinical studies might not be warranted [83].

Alternative models including transgenic rodents (knockout mouse) can be used in some cases but concern exists that these models may not accurately predict the effects of the candidate on pregnancy due to target redundancy or embryo lethality. Rodent homologous proteins that are functionally and structurally analogous to the clinical candidate and recognizes the homologous therapeutic target in the rodent may be a viable alternative to testing of the candidate molecule. However, differences in production and formulation from the clinical candidate may introduce confounding variables, and this approach essentially requires development of a second, parallel drug. Even a structurally dissimilar tool molecule may be used to evaluate on-target effects if functional similarity in the test species is established [40].

The species-specific profile of embryo–fetal exposure during gestation should be considered in interpreting studies. High molecular weight proteins (>5000 Da) do not cross the placenta by simple diffusion. For monoclonal antibodies with molecular weight as high as 150,000 Da, there exists a specific transport mechanism, the neonatal Fc receptor (FcRn) which determines fetal exposure and varies across species [58].

7 Reproductive and Developmental Toxicity Evaluations in the Nonhuman Primate

7.1 Biopharmaceuticals

Nonhuman primates (NHP) are most frequently used for Developmental and Reproductive Toxicity (DART) testing when rodents and/or rabbits are not pharmacologically relevant species. Demonstration of pharmacological relevance is an absolute requirement for biopharmaceuticals (ICH S6: Preclinical Safety Evaluation of Biotechnology-derived Pharmaceuticals). Because

many biopharmaceuticals are human proteins that have been designed to be highly specific for their human target, they often exhibit pharmacological activity in humans and NHPs only. The NHP is often the only pharmacologically relevant species for DART testing. If the biopharmaceutical is less potent in the animal species than in humans, it may still be possible to use that species for DART testing if the dose levels and/or the dosing frequency can be increased sufficiently to produce adequate clinical safety margins taking into consideration the difference in potency. Because of the necessity to use NHPs for some biopharmaceuticals, a larger proportion of biotechnology-derived products have been tested in NHPs than traditional small-molecule pharmaceuticals [39].

7.2 ICH Considerations for Nonhuman Primate DART Studies

The ICH S5(R2) guidance document is applicable to both small-molecule pharmaceuticals and to large-molecule biopharmaceuticals. However, nonclinical development of biopharmaceuticals, as noted in the previous section requires investigators to give consideration to species specificity and immunogenicity as outlined in the ICH S6 guidance document. Human proteins are foreign to animals, and therefore antibodies towards the biopharmaceutical can occur [92, 93]. The formation of anti-drug antibodies (ADA) can lead to enhanced clearance of the biopharmaceutical, reducing exposure which may lead to an overestimation of safety or may over-estimate the toxicity [39]. ADA responses in animals are not expected to be predictive of human responses because in humans the human protein is not xenogenic. Immunogenicity that limits dosing in toxicology studies occurs more frequently in rodents than in NHPs, and therefore the decision to use NHPs for DART studies may be based upon immunogenicity considerations as well as pharmacology [39].

There are many advantages of using NHPs for DART testing. The NHP species most frequently used for general toxicity and for DART testing is the cynomolgus macaque (*Macaca fascicularis*) [94–96]. Cynomolgus macaques are readily available and are not seasonal breeders, so DART studies can be initiated at any time throughout the year. Historical databases are available in this species. Rhesus macaques (*Macaca mulatta*) or marmosets (*Callithrix jacchus*) can also be used although there are certain limitations of using these species, such as seasonal breeding in rhesus macaques [95–97]. Baboons and chimpanzees are not acceptable species for DART testing because of their lack of availability, lack of historical control data, and because of the ethical considerations associated with the use of these species [95].

There are some disadvantages of using macaques for DART testing versus using rodents. Macaques, like humans, have a low natural fertility rate, high spontaneous abortion rate, and are susceptible to prenatal and postnatal losses [58, 96]. Due to single

births and the low fertility rates a large number of NHPs need to be mated in order to produce the desired number of successful pregnancies for DART evaluation [93, 95].

7.3 Study of Fertility and Early Embryonic Development to Implantation Evaluation

The equivalent ICH S5 (R2) Stages A and B cannot really be assessed in NHPs for several reasons. In order to assess a statistically significant number animals in a NHP "fertility" study an evaluation of least 90 NHPs per treatment group would be required to ensure a sufficient number of pregnancies to detect a treatment-related change [58]. The conception rate also is low (45–50 %) therefore assessing the effects on reproduction and fertility in the NHP is not feasible for ethical, logistical, or economic reasons [93, 96]. Indirect measurements of fertility are assessed in NHPs either as separate studies or during repeated dose toxicology studies if sexually mature animals are included. There are disadvantages of using macaques for DART testing. Macaques, like humans, have a low natural fertility rate, high spontaneous abortion rate, and are susceptible to prenatal and postnatal losses [95, 96, 98, 99]. It is both expensive and problematic that a large number of NHPs need to be mated in order to produce the desired number of successful pregnancies for DART evaluation. The M3(R2) guidance document does not differentiate between fertility studies conducted in rodents and fertility (reproduction) studies conducted in NHPs, even though the ICH S5(R2) does not require a non-rodent evaluation of fertility. Therefore, both stand-alone NHP reproduction studies for biopharmaceuticals intended for chronic use in patients of reproductive age, or evaluation of reproductive endpoints as part of chronic toxicology studies (as long as sexually mature animals are used), have been used [58].

According to the ICH M3(R2) guidance document, an evaluation of fertility is needed before the initiation of large-scale or long-duration clinical trials (e.g., Phase III trials). However, ICH S6(R1) notes that if the NHP is the only pharmacologically relevant species for assessing potential effects on mating, conception and implantation the male and/or female reproductive toxicology endpoints can be evaluated in the NHP, either in a stand-alone study design (if there is cause for concern) or (more commonly) incorporated into chronic toxicity studies of adequately long dose duration [86]. If a study is required a sufficient number of sexually mature NHPs need to be sexually mature at the start of the study. Unlike fertility studies in rodents, in NHPs it is not practical to conduct these studies by mating trials; instead, reproduction is evaluated by a number of surrogate markers [39].

When the NHP is the only relevant species, the potential for effects on male and female reproduction can be assessed in repeat dose toxicity studies of at least 3 months duration using sexually mature NHPs. As with rodents body weight and age are the most accurate markers for sexual maturity. A study design would include

standard reproductive organ weights and histopathology, and in females assessment of menstrual cycles. If there is a specific cause for concern based on findings in earlier repeat-dose toxicology studies or based on the target biology, specialized assessments may be included such as sperm count, sperm morphology/motility, testicular volume, spermatogenic staging, and male or female reproductive hormone levels. Alternatively, if a cause for concern exists, a stand-alone male and/or female reproductive study can be conducted. In either case, a weight-of-evidence approach should be used for data analysis, including all available data related to reproductive function [58].

Sexual maturity should be determined in both sexes prior to assignment to a fertility assessment. For both genders, maturity should be confirmed by biological endpoints (e.g., menstrual cycling for females, presence of sperm in semen for males). Body weight is highly correlated with attainment of sexual maturity along with biologic age. Cynomolgus macaques are readily available and are not seasonal breeders, so DART studies can be initiated at any time throughout the year. Historical databases are available in this species. Females should be at least 3 years old and approximately 2.5 kg or more with confirmation of three regular menstrual cycles prior to placement on study [78]. Vaginal swabs are commonly used to assess stage of menstrual cycle by observation of the amount of vaginal bleeding [39, 78]. An important study design consideration in female reproductive testing in NHPs is whether the study is scheduled with respect to the stage of the menstrual cycles of the females (e.g., dosing, blood collections for hormone analyses, and/or necropsy can be scheduled to occur at the same stage of the menstrual cycle) or not. Such adjustment, e.g., start of dosing on day 1 of the cycle, provides more complete and easier-to-interpret reproductive hormone data, and allows all of the females to undergo necropsy at the same stage of their cycles (not doing so sometimes creates apparent test article–related effects on standard toxicity studies). This adjustment to stage of the menstrual cycle can be employed in a stand-alone reproductive study design, but is less practical in the context of a chronic toxicity study [39]. Males reach sexual maturity later (4 plus years) and are also heavier (3.5 kg or greater). Assessment of maturity in males should be confirmed by evaluation of sperm [39]. Males should be dosed a minimum of 60 days prior to mating to cover an entire spermatogenic cycle [78].

In the fertility study, mating and conception are not practical end points in NHPs, the assessment of fertility is focused on effects on reproductive potential rather than fertility per se. The tests conducted in NHPs are essentially the same tests as those conducted in human fertility clinics, and therefore the primate study designs deviate from the rodent study designs in this respect [58].

Potential effects on male and female fertility can be assessed by histopathologic evaluation, assessment of menstrual cycling and

sperm assessments including evaluation of motility, morphology, sperm counts, and testicular size. Hormone analysis: estrogen and progesterone at minimum, since these are the main determinants of the primate ovarian cycle. If effects on estrogen and/or progesterone are detected, additional follow-up analyses may include LH (luteinizing hormone) and FSH (follicle-stimulating hormone) [39]. For stand-alone study designs, an example blood sampling schedule would be every 2 days during the follicular phase (to ensure capture of the estrogen and LH peaks), and every 3 days during the luteal phase. For designs incorporated into a chronic toxicity study, an example blood sampling schedule would be three times per week for approximately 6 weeks (to cover long-duration cycle lengths), conducted once during each phase of the study (pretreatment, dosing, and recovery). Measurement of inhibin A and B to evaluate luteal and follicular phase can also be considered [39].

7.4 Embryo–Fetal Development (EFD) Study

For the EFD NHP study design, sexually mature males (approximately 3–6 years old for cynomolgus monkeys) are used (only for breeding in EFD studies). Sexually mature females (approximately 2–5 years old) are mated to males. For both genders, maturity should be confirmed by biological endpoints (e.g., menstrual cycling for females confirmed by vaginal swabs, presence of sperm in semen for males). This is critical to ensure regular menstrual cycles, predicting optimal time for mating, and maximizing pregnancy rates. Each female is cohoused with a breeder male for 3 days for mating, with pregnancy confirmed by ultrasound on GD18–20; the second day of mating is considered GD0. Based on typical abortion rate of 15–20 % over the entire length of gestation, assignment of 12–14 pregnancies per group should be adequate in most cases (in the absence of test article-related effects) to obtain at least ten viable fetuses per group at C-section [19, 39].

For developmental toxicity studies in NHPs, it is necessary to confirm that the animals are pregnant by the presence of a gestational sac and/or yolk sac, before dosing is initiated (usually by GD18–20) [58]. A standard NHP embryo–fetal development protocol has been developed in which the pregnant females are dosed during the period of organogenesis (GD20–50) and fetuses are obtained by caesarean section (C-section) at a time when adequate morphological and skeletal development (GD100) can be assessed [39, 100]. The basic study design is modeled on the rodent and rabbit study designs. However, experience with NHP embryo–fetal development studies and an understanding of human biology has indicated that this basic study design is not always the optimal design for all types of molecules or for all therapeutic classes of molecules. However, it is becoming clear that when NHPs are being used for DART testing, a standardized segmented approach may not be optimal from an animal-usage or scientific basis [39]. A single well-designed developmental study that takes into

consideration the type of molecule being tested, the pharmacology, and the intended clinical use may be preferable. For example, a single enhanced prenatal and postnatal development (ePPND) design option may be sufficient to evaluate developmental effects of therapeutic monoclonal antibodies [58]. In this enhanced study design, NHPs would be dosed from the beginning of organogenesis (GD20) through to birth, and the offspring would be evaluated for a number of months postpartum for any functional or structural defects [58].

Some biopharmaceuticals (e.g., monoclonal antibodies), transport across the placenta may not occur until the second or third trimester [58]. Also, potential target organ systems such as the immune system are likely to continue development during this time [96]. For macaques, the conventional dose duration of GD20–50 may be extended for some pharmaceutical and biopharmaceutical EFD studies to optimize exposure of the fetus and later-developing organ systems for example the immune system. Dosing may be extended until GD90–100 (or later) and C-sections delayed until GD120–140 (natural birth is typically ~GD160). C-sections could be performed even later in gestation (e.g., GD150). However, macaques are prone to late pregnancy losses such as stillbirths or premature births, and that will reduce the number of fetuses available for examination [39, 99]. Therefore, a larger number of pregnancies are needed in EFD studies that are carried to near term, in order to ensure a sufficient number of fetuses for examination.

Standard design (e.g., small molecule) is a control group and three dose groups (test article treated). Biologics will frequently employ a control group and two dose groups, usually multiples of the clinical dose if no toxicity is expected. The use of a control group and a single treatment group has also been suggested to reduce NHP use [19, 83, 101, 102]. Dose selection is usually based on data from general toxicity (nonpregnant) studies, but a dose-ranging study in pregnant females may be conducted (e.g., $n = 5$ per group) [39].

Since the fetal immune system continues to develop until later in gestation dosing may be extended into the fetal period for test articles with known effects on the immune system [103]. The same logic may apply to other organ systems continuing development such as the nervous system. For monoclonal antibodies, based structurally on an IgG backbone, dosing is best extended into the fetal period, to GD90 or 100 or even later to ensure fetal exposure [58, 94]. This is because Fc–receptor-mediated placental transfer of IgG occurs during the fetal period, especially the late fetal period, in both humans and NHPs [104]. GD140 is suggested as the maximum practical time point for termination of dosing and C-section, to avoid increasing probability (and confounding factor) of live birth and stillbirths after this interval [39].

Dosing may be extended until GD90–100 (or later) and C-sections delayed until GD120–140 (natural birth is typically ~GD160). C-sections could be performed even later in gestation (e.g., GD150). However, macaques are prone to late pregnancy losses such as stillbirths or premature births, and that will reduce the number of fetuses available for examination [99]. Therefore, a larger number of pregnancies is needed in EFD studies that are carried to near term, in order to ensure a sufficient number of fetuses for examination.

Ultrasounds are conducted throughout gestation until scheduled C-section as a check of fetal viability. Can also include a number of fetal measures [e.g., fetal length, long bone length, abdominal circumference, head circumference, biparietal diameter, and/or occipitofrontal diameter although these may be considered optional for an EFD study due to measurements conducted at C-section [105]]. GD140 suggested as maximum practical for termination of dosing and C-section, to avoid increasing probability (and confounding factor) of premature births and stillbirths. Evaluations immediately after C-section include evaluations of amniotic fluid (volume, clarity), placenta (weight, number of disks, disk(s) diameter, appearance of maternal and fetal surfaces, location of cord insertion), and the umbilical cord (length) [105].

Maternal toxicokinetics (TK) can be conducted to evaluate maternal/fetal exposure ratios and may help determine if TK in pregnant females is different from that determined in nonpregnant animals in general toxicity studies. For biologics, determination of anti-drug antibodies is usually included to help determine if the test article is immunogenic and to consider potential impact on TK [78]. Clinical pathology can include hematology, serum chemistry, coagulation parameters, and/or urinalysis can be included if a specific purpose only [e.g., documentation of expected maternal toxicity, confirmation of expected pharmacodynamic (PD) effect] is expected. Maternal immunologic evaluations: e.g., immunoglobulin levels, (immunophenotyping) can add significantly to the assessment of maternal toxicity and response to the biopharmaceutical under evaluation [104].

External examination of fetus includes fetal weight and morphometric measurements (e.g., crown-rump length, head circumference, long bone length, and anogenital distance). Visceral examination of fetus for internal organ morphology can also include weighing selected tissues/organs. Selected fetal tissues may also be evaluated by histopathology and/or immunohistopathology. GD100 and 140/150 fetuses can be processed and stained with Alizarin red for skeletal evaluation. Alternatively, if C-sections are conducted late in gestation, it may be more convenient to take radiographs for the skeletal exam (avoids prolonged time required to process larger fetuses for Alizarin red) [39].

Umbilical cord blood can be collected for TK evaluations. Blood sampling allows confirmation of fetal exposure and calculation of maternal/fetal ratios. Evaluation of anti-drug antibodies usually included for biologics. Levels of test article in amniotic fluid can also be determined. Since the number of NHP fetuses available for gross visceral examinations is relatively small, histopathology of major organs or target organs can also be included. Evidence of histologic effects may be of help in the consideration of a mechanism of abnormal development of some organs or tissues. Evaluations of the fetal immune system can include: flow cytometry (limited by blood volume), lymphoid organ weights (mainly spleen due to small size of fetuses), histopathology, and/or immunohistopathology [39, 104].

7.5 Enhanced Study Design for Prenatal and Postnatal Development (e-PPND)

Prenatal and postnatal development study protocols in NHPs are less well established than the embryo–fetal development protocol. The enhanced prenatal and postnatal development study (ePPND) is a study that combines endpoints of both the traditional EFD and PPND studies in which dosing is extended throughout gestation to parturition [106]. The study design which appears to be the most appropriate to evaluate mAbs assesses the indirect effects in early gestation and the effects of fetal exposure and through postnatal development [39].

Developmental toxicity studies in NHPs can only provide hazard identification. The sponsor should justify the study design if other NHP species are used. The developmental toxicity studies in NHPs as outlined above are just hazard identification studies; therefore it might be possible to conduct these studies using a control group and one dose group, provided there is a scientific justification for the dose level selected [19]. An example of an appropriate scientific justification would be a monoclonal antibody which binds a soluble target with a clinical dosing regimen intended to saturate target binding. If such a saturation of target binding can be demonstrated in the animal species selected and there is an up to tenfold exposure multiple over therapeutic drug levels, a single dose level and control group would provide adequate evidence of hazard to embryo–fetal development [83].

If women of child-bearing potential are included in clinical trials prior to acquiring information on effects on embryo–fetal development, appropriate clinical risk management is appropriate, such as use of highly effective methods of contraception (see ICH M3(R2) Guideline). For biopharmaceuticals pharmacologically active only in NHPs, where there are sufficient precautions to prevent pregnancy (see ICH M3(R2) Guideline, Section 11.3, Paragraph 2), an EFD or ePPND study can be conducted during Phase III, and the report submitted at the time of marketing application [83]. When a sponsor cannot take sufficient precaution to prevent pregnancy in clinical trials, either a complete report of an

EFD study or an interim report of an ePPND study should be submitted before initiation of Phase III. Where the product is pharmacologically active only in NHPs and its mechanism of action raises serious concern for embryo–fetal development, the label should reflect the concern without warranting a developmental toxicity study in NHPs and therefore administration to women of child-bearing potential should be avoided [83].

For the single ePPND study design described above, no Caesarian section group is warranted, but assessment of pregnancy outcome at natural delivery should be performed. This study should also evaluate offspring viability, external malformations, skeletal effects (e.g., by X-ray), and, ultimately, visceral morphology at necropsy. Ultrasound is useful to track maintenance of pregnancy but is not appropriate for detecting malformations. These latter data are derived from postpartum observations. Because of confounding effects on maternal care of offspring, dosing of the mother postpartum is generally not recommended. Other endpoints in the offspring can also be evaluated if relevant for the pharmacological activity. The duration of the postnatal phase will be dependent on which additional endpoints are considered relevant based on mechanism of action [83].

Pregnancies are allowed to continue to natural delivery, and the infants are subsequently evaluated for growth and development, most commonly up to 6 months. However, the postpartum duration can be as long as 12 months of age, if for example there is prolonged recovery from pharmacodynamic effects in the infants or if CNS assessments involving memory and learning are part of the study design. Hence, in this study design ultrasound measurements during pregnancy confirm pregnancy status and the external, visceral, and skeletal evaluations are done on the infant [39, 83]. The advantage of this study design is the reduced use of NHPs by combining the EFD with the PPND study. The disadvantage is that if there is an increase in spontaneous abortions during the early or late stage of gestation or if the test article is an abortifacient, then a large number of animals may be needed to ensure a sufficient number of neonates for evaluation [39, 106].

7.6 Juvenile Toxicity Studies in NHPs

The goal to develop a nonclinical development plan that adequately addresses safety concerns for the intended use in specific pediatric populations must take into consideration the specific pharmacologic properties of the biopharmaceutical, the clinical development plan, the therapeutic indication and age of pediatric population. Both EMA and FDA emphasize the importance of the case by case approach in developing the study design to support the clinical program and clinical use in pediatric patients. Design of the Juvenile Toxicity Study in the NHP must be designed taking into consideration the existing nonclinical toxicology data available

Table 2
Species considerations for juvenile toxicity testing with biopharmaceuticals

Species	Advantages	Disadvantages
Nonhuman primate	Maybe the only pharmacologically responsive model for testing	• Large vendor colonies needed to procure adequate numbers of appropriately aged animals, of both genders
	• Information from previous General Toxicology and/or DART studies maybe available from nonclinical development program	• Vendors typically do not have animals available (approx. 12 months old) to provide match to clinical pediatric age range.
	• Size and body weight allows collection of multiple biologic specimens and evaluation of a variety of endpoints	• Long development period until sexual maturity prohibits evaluation of full span of postnatal development.
	• Postnatal developmental time-points of many organ systems are well characterized	• Limited historical control data for some data types due to infrequent use for juvenile testing
	• Immunoassays available	• Potential animal welfare issues due to separation from mother
	• Techniques are available in adult NHPs that have been adapted to neonatal and juvenile age ranges	• Expensive due to animal costs, time required (vs. rodent)
	• Standardized tests are widely used for neurobehavioral testing, including learning and memory	• A limited number of laboratories have experience and technical expertise to conduct these studies

Adapted from Morford [39]

and because each molecule is different (Table 2). Extensive interactions with regulatory agencies are critical for the development of the juvenile toxicity study design to obtain their concurrence on the study design. A targeted approach for the NHP study design should be used to specifically address toxicity concerns uncovered in previously conducted toxicity studies. The study should be targeted for a detailed evaluation focusing on specific organ system(s) of concern [39].

NHPs can be a useful model for juvenile toxicity testing, and may be the species of choice when used previously for general toxicity and/or reproductive toxicity testing in a drug development program. Their obvious phylogenetic similarity to humans, NHPs have advantages compared to other species as relates to evaluation of the developing immune system, nervous system (including behavior), and skeletal system. Consensus on the age for animals in juvenile NHP studies has not been reached; however, in general an age range between 9 and 36 months is appropriate and practical. This age spans the gap between the neonatal/infant phase of developmental and reproductive toxicology studies and the age typically used for general toxicology studies (young adults but sexually

immature). The design of a NHP juvenile toxicity should be developed on a case-by-case basis given the consideration of the pharmacologic properties of the test article and the clinical situation being supported [39].

Awareness of model-related constraints of NHP consists of animal availability and study logistics. It is pivotal to distinguish between neonatal and juvenile (i.e., before sexual maturity) animals. Pregnant NHPs are typically not available from breeders. Consequently, in cases where neonatal animals are needed, the need for breeding at the site of study conduct must be considered. In reality, a study utilizing neonatal monkeys of this age would require breeding the animals at the Testing Facility to obtain neonates that could then be dosed directly. This has significant implications on the prestudy timing as average pregnancy duration in cynomolgus monkeys is 160 days and gender distribution of neonates would be random. The inability to control gender distribution has practical implications as well as animal welfare implications. Imbalances in the number of male and female infants will affect study design and, if too many neonates of one gender are being delivered. Also, depending on the specific questions being addressed by the juvenile toxicity study (such as concerns regarding male reproductive organs), only a single gender may be needed. Transport regulations may also have a bearing on the conduct of NHP juvenile toxicity studies in terms of the age range of available animals. International regulations for shipping of juvenile NHP (without their mothers) generally require the animals to be at least 12 months old. Domestic country-specific shipping regulations may allow juvenile monkeys to be shipped from vendors at approximately 8–9 months of age (i.e., postweaning). Younger animals may be available but must be shipped along with their mothers, and as NHP breeders usually prefer not to provide maternal animals along with infants, the conduct of studies in animals younger than 9 months poses a significant challenge [107].

As a general recommendation compromising between animal availability and study feasibility, an age range of 9–36 months has been suggested for NHP juvenile toxicity studies. Given the recent changes in NHP transport regulations, 12–36 months may become the reality. When close to 36 months age range, it is important to evaluate whether some animals have already reached sexual maturity, as may be the case for female cynomolgus monkeys but less likely for male animals [39, 93].

There are additional considerations for conducting juvenile NHP studies. Many biopharmaceuticals induce immunomodulatory activity and consideration should be given in these studies to evaluate immune system function. Due to the length of the maturation period in NHPs, assessment of reproductive capacity is not feasible. A clear delineation of the stage of reproductive development (immature, peripubertal, or mature) should be made for individual animals

at the time of histopathologic exam and these data should be used in the assessment of potential juvenile toxicity, Additionally, NHPs growth and development is difficult to monitor in juvenile macaques and is further complicated by individual variability, protracted growth period and the long period of skeletal development (7–8 years well beyond a feasible observation period) [19]. Historical data is employed to assess developmental milestones and skeletal development can be assessed by radiography and DEXA [107].

One of the current challenges in drug development for pediatric populations is determining when to initiate a juvenile toxicity study. In Europe, the EMA generally expects that a PIP be initiated once adult PK data are available, which is usually early in clinical development. This request for a relatively early pediatric strategy is not a current part of regulatory expectations in the USA (FDA). For companies planning a worldwide development strategy, this presents challenges to develop a concurrent strategy that will have global regulatory acceptance [107].

8 Alternative Methods Used for Reproductive and Developmental Toxicity Testing

Alternatives to animal testing in developmental toxicology have been the subject of at least three decades of research. These investigations have a common goal to reduce animal experimentation, to refine effect assessment and mechanistic studies, and to accelerate and simplify safety testing in an area of toxicology that uses relatively many animals. Many alternatives have been developed over the years with different complexities, using biologic material ranging from continuous cell lines to complete embryos. The validation of alternatives and their application in testing strategies is still in its infancy, although significant steps towards these aims are currently being made. The introduction of the genomics technology is a promising emerging area in developmental toxicity testing in vitro. Future application of alternatives in testing strategies for developmental toxicity may significantly gain from the inclusion of gene expression studies, given the unique program of gene expression changes in embryonic and fetal development. Existing alternative methods for developmental toxicity testing can roughly be subdivided into three types of systems with increasing complexity: cell cultures, organ cultures, and embryo cultures. Cell cultures have the disadvantage of simplicity with the advantages of ease of performance and minimal or no animal use, while on the other end of the spectrum, embryo cultures are advantageous in that they incorporate more developmental mechanisms and end-points with the disadvantages that they are more laborious and use more animals [108].

Capabilities for early screening of drug candidates for potential developmental toxicity have evolved quickly in the last several years

and several assays are becoming a regular part of the extensive process of selection of optimal molecules to advance for development programs for potential medicines. Developmental toxicity assays selected for use must have high throughput capabilities and deliver data quickly with precision and reliability to enable differentiation [prioritization] of compounds. Test systems, such as whole embryo culture, zebra fish and stem cell cultures, whose developmental responses are well characterized have the potential to be used in combination with other screening assays to generate data that allow for comparative analysis of toxic potency. More often, however, the data from early screening alerts a company to a potential developmental risk. In such cases, the subsequent developmental studies can be modified, or focused, or mechanistic studies can be devised, to better assess that potential risk. These assays can aid in compound selection early in the drug development process but they will not replace the more thorough regulatory testing that must occur for a drug to be approved for use [109].

The main issues of the validity of a toxicity test system are its reproducibility and its predictivity. As to reproducibility, this pertains not only to results within one laboratory, but in addition the test should give comparable results in different laboratories, preferably allowing some flexibility in methodology. Flexibility in methodology facilitates the introduction of a test system in a laboratory and when possible is a sign of robustness. The ECVAM validation study of rodent postimplantation embryo culture allowed the use of different rat strains, different culture media and different culture apparatus. Nevertheless, a high reproducibility of results between laboratories was found [110].

Whole embryo culture systems offer the most complete in vitro alternatives for animal testing. The embryo culture which has received most attention is undoubtedly rodent postimplantation whole embryo culture. There are currently three validated alternatives in developmental toxicology for which standard protocols are available from the website of the European Centre for the Validation of Alternative Methods (ECVAM). These are the embryonic stem cell test (EST), the limb bud micromass, and the rat postimplantation whole embryo culture (WEC). These three tests have been formally validated in a study involving four independent laboratories for each alternative test system and 20 chemicals that were tested in a double-blind protocol in each test. The application of these tests as valid alternatives to animal testing in developmental toxicology is still under discussion [108, 110, 111].

The toxic effects of exposures to xenobiotic compounds may become apparent either directly after exposure or after a longer period of time. Especially in the latter situation, short-term in vitro assays may not be able to detect these effects at the level of morphologic events. However, at the molecular level, early effects of test compounds may be detectable that predict toxic

effects that may occur in due time later on in development in the in vivo situation. The application of genomics and proteomics to determine early effects of compounds in in vitro assays may therefore increase the sensitivity of in vitro assays and bridge the gap between early markers and late effects [108].

The application of differential gene expression analysis as a tool for studying the toxicity of chemicals in in vitro systems looks especially promising in the field of developmental toxicology. Embryo–fetal development is uniquely driven by gene expression changes which are programmed in time and space. Interference with programmed gene expression is likely an early effect on the route to malformation. A wealth of model systems employed in developmental biology including knock-out mutants and morpholino knockdowns in the Zebrafish model [112].

8.1 Embryonic Stem Cell Test (EST)

The embryonic stem cell test (EST) described by the European Centre for Validation of Alternative Methods (ECVAM) used extensively for in vitro screening for possible embryotoxic effects of chemicals and drug candidates. Mouse embryonic stem cells (mESC) are maintained in culture as pluripotent cells by incubation with leukemia inhibitory factor (LIF). Cytotoxicity is determined for 3T3-A31 fibroblasts and for ES cells. The test evaluates the ability of the ES cells to differentiate into functional contracting cardiomyocytes after exposure to the test article. The stem cells can be further evaluated for the ability to differentiate in vitro, to all of the components of the embryo. The processes involved in establishing each embryonic layer (e.g., ectoderm, mesoderm, and endoderm) and the subsequent differentiation of these embryonic cells are recapitulated in this model. Thus, the model has the capability to assess many of the events associated with embryogenesis. One of the future advances will likely be the use of molecular markers to evaluate phenotypic differentiation. Another advantage of this system is that it carries out a direct comparison between differentiation and cytotoxicity/proliferation. This comparison may add to our ultimate characterization of xenobiotics as developmental toxicants in vivo. Because of the nature of the ESCs, the EST is amenable to relatively high-throughput modifications (robotics, scale reductions) for culture and a point-by-point visual evaluation of a physical structure to facilitate morphologic evaluations (such as contraction or large lipid droplets) [113, 114]. Also, once derived, the mESC do not require the use of animals, which is a major benefit in some contexts. In contrast, for several human ESC lines maintenance culture of the pluripotent cells does use a feeder layer of mouse embryonic fibroblasts that requires the use of additional animals. Advances are currently being made for xeno-free hESC culture that would not require animals [113–115].

Three endpoints are commonly evaluated to assess the embryotoxic potential of the test article:

1. Inhibition of differentiation into beating cardiomyocytes
2. Outgrowths compared to cytotoxic effects on the stem cells
3. Differentiation of the 3T3 fibroblasts

An extensive amount of research has been conducted with variations of this test to validate the methods using 20 test compounds of known embryonic toxicities [113, 114]. Classification of teratogenicity was based on the premise that a threshold of metabolic perturbation could be identified for individual metabolites that is associated with developmental toxicity. This threshold of metabolic change is called the teratogenicity threshold and is a measure of the magnitude of metabolic perturbation required to differentiate teratogens from nonteratogens. The teratogenicity threshold was empirically generated for ornithine, cystine, and the o/c ratio by iteration through a range from 10 to 25 % change to identify a one- or two-sided asymmetric threshold that was able to classify the training set with the greatest accuracy and highest sensitivity. In the case of a tie in classification accuracy and sensitivity between one- and two-sided thresholds, one-sided thresholds were given priority to favor simplicity. A teratogenicity threshold was determined for each phase of the study, since the assays performed in Phase 1 used only a single concentration of each compound and the targeted biomarker assay developed in Phase 2 utilized an exposure-based approach. The teratogenicity threshold was determined in Phase 2 using only the results from the training set. This threshold was then applied to the results from the test and application sets [113–115].

The stem cell test is relatively simple to carry out, and the main endpoint (scoring of beating) requires no in-depth knowledge of phenotype and morphologic development. The assessment of proliferation and cytotoxicity relies on standard, well established assays already in use in many laboratories. Thus, there is only limited additional training required for laboratories already adept at cell culture to carry out the EST as described in the ECVAM protocol. The assay is more amenable to relatively high-throughput modifications (robotics, scale reductions). The EST seems to work best at both ends of the activity spectrum: if a compound is classified as not a developmental toxicant, there is a very high chance (>90 %) that this will be true. For strong developmental toxicants, the assay finds them all, but the assay can also classify some weak and nontoxicants as strong developmental toxicants. The current predictive model for the EST depends largely on the absolute concentration for the various toxicities; i.e., a very low effective concentration for producing cytotoxicity will drive the model to categorize the compounds as "Strong," even if its in vivo activity is weak or less. And,

like the WEC, this assay lacks a maternal component, and has limited metabolic capability (although the use of S9 can add this when necessary); this is both a strength and a weakness. The strength is that metabolites and parent compounds may be tested individually and the true active agent identified. Conversely, identifying and obtaining the metabolite(s) may be a challenge [114, 115].

There are disadvantages of the EST as well. The absence of a meaningful maternal component means that the assay is currently limited in its ability to model the direct effects of the compound on the developing system and predict dose-limiting maternal toxicity. It is not impossible to predict developmental effects produced by changes in maternal physiology (e.g., acidosis) that, in turn, alter development in vivo. The current EST model over-predicts some activities, such that many developmental non-toxicants are classified as toxicants by the assay. Effectively segregating nontoxic and weakly toxic compounds is one of the greatest challenges for the assay. Another potential liability of the EST is its reliance on one differentiation outcome in the assessment. The identification of beating cells as a marker of cardiomyocyte differentiation may be confounded by an effect of the compound (test agent) on the contraction of the cardiomyocytes (such as altered energy production) or a direct cardiomyocytes toxicant. Also, there is no difference in the assessment of differentiation when one well has ten beating cells compared with a well that contains 10,000 beating cells. Thus, this lack of discrimination in what constitutes "differentiation" may add to a lack of specificity in the assay. Further, the random differentiation in embryoid bodies and fact that the differentiated cells produce yet unknown growth/protective factors and cell types adds another layer of uncertainty. In general this assay also does not produce late-differentiating cell types that can be seen in, e.g., teratomas [114].

An exciting area under development is the use of molecular markers to assess multiple differentiation phenotypes in mESC after undirected differentiation. Using a quantitative approach, the relative level of mRNA for specific molecular markers (e.g., α-MHC for cardiomyocytes) can be assessed to determine differentiation to each phenotype and the relative differentiation to multiple phenotypes (e.g., ectoderm compared with mesoderm). Using lineage specific markers may also aid in the prediction of a target-tissue effect of a test agents as a developmental toxicant [114].

8.2 Whole Embryo Culture (WEC)

Whole Embryo Culture (WEC) techniques have a long history in evaluation of toxicity to the developing rodent embryo. The assay is run by explanting rodent embryos with their yolk sacs on approximately gestation day (GD) 9.5 or 10 for rats. The conceptus is then cultured on a rotating platform in a mixture of serum and culture medium (and test article, if applicable) for 44–48 h with increasing proportions of oxygen added in the gas overlayer. At the end of the

culture period, the conceptus is evaluated for the degree of maturity of various endpoints, e.g., number of somites, optic development, forelimb development, neural tube development, etc. Each of these is given a score (morphologic score based on methods developed by Brown and Fabro; Klug et al.; Van Maele-Fabry et al. [116–118]), and the scores are summed for each concentration tested. The presence and type of any malformations are also noted. The readout of the assay can be: (1) the concentration at which malformations begin to be evident; (2) a comparison across different compounds of the types and severity of malformations seen at a given concentration; or (3), the results from the predictive linear discriminant analytic formulae when using the ECVAM version [117–119].

Rodent WEC was included in a validation study of embryotoxicity tests conducted by the ECVAM [110]. The reproducibility of the WEC test as well as the concordance between the embryotoxic potential derived from the in vitro data and from in vivo data were good according to predefined performance criteria. The prediction model correctly classified 80 % of the 20 tested compounds for all embryotoxicity classes (non-embryotoxic, weakly embryotoxic, and strongly embryotoxic). More information on the validation study, including comprehensive protocols of the methodology, is available on the ECVAM web site [114]. Rabbit embryos can be cultured relatively easily using techniques similar to those developed for rat and mouse WEC. Typically, one rabbit embryo is allocated per test substance concentration, with continuous flow gassing using a rotating incubator. Endpoints evaluated are viability, growth, morphologic scoring [based on the rat WEC method of Brown and Fabro as modified for rabbit embryos by Carney et al. [119]], and biochemical measures (total protein, DNA). One key difference between rat and rabbit embryos concerns the yolk sac: GD 9 rat embryos rely on histiotrophic nutrition through an inverted visceral yolk sac in which the embryo is enclosed within the yolk sac; rabbit embryos do not have an inverted yolk sac, and lie outside the yolk sac until approximately GD 13. However, yolk sac-mediated histiotrophic nutrition is specific to rodents, and may be less relevant to humans (that rely on hemotrophic nutrition). For the future, as with rat and mouse WEC, it is anticipated that functional endpoints (endocytosis, proteolysis), gene expression (yolk sac transporters), and imaging techniques (Micro CT, magnetic resonance imaging, morphometry of embryonic volume, or specific landmarks of development) will become useful endpoints for rabbit WEC analysis. Rabbit WEC can also help address species differences in developmental toxicity responses [114].

For many labs, growth parameters (e.g., crown–rump length, protein content) are good predictors of potential embryotoxicity with the added advantage of being continuous variables (unlike morphology-based parameters). Accordingly, it is proposed to change the criteria used to evaluate embryotoxicity in WEC so as not to rely on morphologic scores alone. Therefore, the IC_{NOAEL}

for a compound may be related to total morphologic score, but there is often a steep dose–response curve resulting in a rapid transition from non-embryotoxic to strongly embryotoxic effects. In these cases, the maximum inhibitory concentration (IC_{MAX}) would be equivalent to the concentration producing the highest malformation incidence. In addition, embryos obtained from different species provide different predictions, as is also true in vivo. WEC should not be used for human risk assessment purposes at the present time, based in part on the absence of a maternal compartment and the very limited exposure window. Compound class-specific prediction models may improve predictivity within pharmaceutical or chemical structural classes [114, 115].

The species used for WEC (mouse, rat, and rabbit), are the same species that are used most commonly in whole animal reproductive toxicity assays, thus allowing a direct correlation between in vitro and in vivo findings, and against mouse embryonic stem cell data. WEC can recapitulate in vivo embryonic development for up to 48 h. Embryos can be treated in a milieu isolated from maternal effects, but some maternal effects can be introduced into the culture system if desired (e.g., obtain culture serum from treated animals, add known maternal metabolites to culture medium, hyperthermia). WEC is useful for mechanistic studies, prioritization/screening of compounds, studying intrinsic differences between species, providing adjunct information for regulatory/risk assessment purposes, or to further investigate in vivo findings to increase confidence in the data while minimizing the use of animals [114, 115].

The WEC cannot replace in vivo developmental toxicity studies at the present time because it does not recapitulate the maternal–fetal interactions or expose the conceptus for the gestational period of concern (implantation to near-term); isolation from maternal influences (metabolism, toxicity) that may contribute to in vivo effects; restriction to a relatively narrow developmental window that may not allow it to capture some manifestations of developmental toxicity; and variation across aliquots or collections of serum. At the present time, no US regulatory agencies (i.e., FDA, EPA) are using WEC for regulatory decisions because WEC has not been validated for that purpose. As a consequence, in vitro alternatives are unlikely to be accepted in the near future [114].

8.3 Zebrafish Embryotoxicity Test (ZET)

A Zebrafish (*Danio rerio*) teratogenicity assay has been developed as a screening tool and has evaluated for its ability to predict the teratogenic potential of chemicals and drug candidates. The results indicate that this assay is promising for screening compounds for teratogenic potential. Zebrafish share most of the signaling systems found in mammalian embryos, although heart development may follow the same pattern in rodents and zebrafish. The developing Zebrafish (*Danio rerio*) is an in vivo developmental model with many of the advantages of in vitro systems and much potential to

meet the needs associated with a simple, inexpensive, and rapid assay to screen for teratogenicity. Zebrafish are inexpensive and easy to maintain, breed, and rise and also their development is similar to that of with many molecular pathways are evolutionarily conserved between zebrafish and humans. It has been shown that zebrafish possess orthologs to the majority (86 %) of human drug targets. Toxicologists have used zebrafish embryos for many years to identify endpoints and elucidate mechanisms of developmental toxicity for a number of chemicals. Zebrafish also offer the unique advantage of testing compound effects during the entire period of organogenesis due to their rapid, external development. This important aspect is not possible to achieve with some other commonly used teratogenicity screening assays such as the Embryonic Stem Cell Test (EST) or rodent Whole Embryo Culture (WEC). In addition, the transparency of the surrounding egg membrane (chorion) and embryonic tissues make it possible to observe development throughout the embryonic period [120].

Recent studies with the "Zebrafish *Danio rerio* Teratogenic assay" (*Dar*T) have demonstrated the general utility of zebrafish to screen compounds for teratogenic potential. A review conducted by Brannen et al. to develop a zebrafish teratogenicity screen that would allow for the characterization of teratogenicity as it relates to specific abnormalities and concentration-response and to evaluate the concordance and predictivity of that assay was conducted on 34 in vivo teratogens and nonteratogens, including many exploratory and marketed pharmaceutical compounds. In the course of this work, a morphological scoring system was developed, and the features of this assessment are also illustrated here. The results demonstrate a high rate of success in predicting in vivo mammalian teratogenicity with zebrafish embryos [120].

In a review of 34 compounds, which were a mix of ECVAM validation compounds and pharmaceutical compounds with characterized in vivo teratogenic potential, were evaluated using the ZET test protocol. The basic assay design involves dechorionated Zebrafish embryos treated with compounds at 24 h post-fertilization (pf), at which time the embryos were actively undergoing organogenesis. A 48 h exposure period is typical and morphology is assessed at ~120 h pf (~4 days pf). Two measurements of general toxicity were evaluated in prediction modeling. The first measurement adapted a practice previously used by ECVAM in predictive modeling of the rodent whole embryo culture and mouse embryonic stem cell assays, where the respective compounds were evaluated in a dose range in NIH3T3 fibroblast cells. The second general toxicity measurement involved evaluating the compounds in a concentration range in zebrafish embryos and determining the general toxicity concentration based on a 25 % lethal dose concentration (LC_{25}). A morphologic scoring system was developed that assessed various structures and organs in the Day 5

pf larvae; this included a score for the severity of the dysmorphology. Predictive model classification of each compound involved calculating a ratio of the general toxicity concentration to the no-observed-adverse-effect concentration (NOAEC) based on gross morphology. The results of this comparative review by Brannen et al. indicates that the cumulative concordance of the prediction model outcome with in vivo teratogenicity data was 92 %, with a 94 % success rate in positively identifying in vivo teratogens and a 86 % success rate in positively identifying in vivo nonteratogens. In addition, there was an 87.5 % success rate in positively characterizing in vivo morphologic outcome (either no adverse effect on fetal morphology or positive identification of at least one affected structure/organ system associated with in vivo exposure to the compound) [120].

Although this work emphasizes the overall utility and predictivity of the zebrafish assay for assessing development toxicity, there are still several unresolved questions regarding the broad use of zebrafish in predictive teratogenicity. Very little is known about the pharmacokinetics and metabolism of waterborne compounds in zebrafish embryos and larvae, and exploring these properties was not a goal of this study. A better understanding of zebrafish exposures and the relevance of pharmacokinetic and metabolic mechanisms to mammals would be quite valuable to improve interpretation of findings in the zebrafish assay in the context of how they relate to teratogenic potential in humans and other mammals. Furthermore, in these experiments, the chorion from each embryo was removed in an attempt to improve compound exposure to the embryo.

As a teratogenic screening tool, zebrafish offers an entire organism and all stages of development, not a conceptus for a limited part of development or isolated cells in cell culture. This intact model allows all the cells and tissue layers of the conceptus to interact normally, and brings completeness unavailable in other models. The small size of the embryos also means that compound requirements are minimal, which is of benefit when testing novel compounds that must be created de novo before testing. Additionally, one can produce an allelic series of hypomorphic embryos, where progressive knock-down of gene expression produces successively more-impacted phenotypes. Early indications suggest that this model has the potential to provide predictivity that is at least as good as existing models, and perhaps better [120].

9 Perspective and Future of Reproductive and Developmental Toxicology

While the science and knowledge base of reproductive and developmental toxicants has progressed dramatically since the thalidomide tragedy the basic protocols and study designs using animal models are little changed since the promulgation of the requirements for these tests for pharmaceutical registrations.

Our understanding of embryonic development, molecular genetics and mechanisms of toxicity has experienced exponential growth. New tools are now available to assess endpoints commonly evaluated in these studies yet we are bound by regulatory requirements as no in vitro methods are to date acceptable substitutes for whole animal mammalian models. The current developmental and reproductive testing paradigm is comprehensive but the cost and number of animals required for these studies is high.

At the present time there are opportunities to incorporate new technologies and approaches to testing into the existing assessment paradigm, or to apply innovative approaches to various aspects of risk assessment. Developmental toxicology testing can be enhanced by the refinement or replacement of traditional in vivo protocols, including through the use of in vitro assays, studies conducted in alternative nonmammalian species, the application of new technologies, and the use of in silico models. Potential benefits to the current regulatory process include the ability to screen large numbers of chemicals quickly, with the commitment of fewer resources than traditional toxicology studies, and to refine the risk assessment process through an enhanced understanding of the mechanisms of developmental toxicity and their relevance to potential human risk. As the testing paradigm evolves, the ability to use developmental toxicology data to meet diverse critical regulatory needs must be retained [109].

The capabilities for early screening of drug candidates for potential developmental toxicity have evolved tremendously in the last several years and such assays are becoming a regular part of the extensive process of selection of optimal molecules to be advanced for registration and use as medicine. The developmental toxicity assays selected for use must have high throughput capabilities and deliver data quickly with precision and reliability to enable the selection process [prioritization] of drug candidates for further development. Test systems, such as whole embryo culture, zebra fish and stem cell cultures, whose developmental responses are well characterized have the potential to be used in combination with other screening assays to generate data that allow for comparative analysis of toxic potency. These data contribute to an overall evaluation of a drug candidate and may eliminate a particular molecule [or class of molecules] from further evaluation. More often, however, the data from early screening alerts a company to a potential developmental risk. In such cases, the subsequent developmental studies can be modified, or focused, or mechanistic studies can be devised, to better assess that potential risk. These assays aid in eliminating many drug candidates in the drug development process they will not replace the regulatory testing that must occur for a drug to be approved for use [109].

References

1. Lenz W (1988) A short history of thalidomide embryopathy. Teratology 38:203–215
2. Zimmer C (March 15, 2010). Answers begin to emerge on how thalidomide caused defects. New York Times. Accessed on March 21, 2010
3. Reversal of fortune: how a vilified drug became a life-saving agent in the "War" against cancer - Onco'Zine - The International Cancer Network (November 30, 2013)
4. Seidman LA, Warren N (June, 2001) Pharmaceutical regulation in the United States: history and a case study. Madison Area Technical College, Madison, WI. lseidman@madison.tec.wi.us
5. McFadyen RE (1976) Thalidomide in America: a brush with tragedy. Clio Med 11 (2):79–93
6. Herbert Burkholz (September-October 1997) Giving Thalidomide a second chance. *FDA Consumer magazine*. Available at http://www.fda.gov/fdac/features/1997/697_thal.html
7. From a speech "Denial of Approval for Thalidomide in the United States", by Frances O. Kelsey, presented at the Medicine and Health Since World War II, National Library of Medicine, Bethesda, MD, Dec. 9, 1993.
8. 50 Years: The Kefauver-Harris Amendments
9. Goldenthal EJ (1966) Guidelines for reproduction studies for safety evaluation of drugs for human use. Drug Review Branch, Division of Toxicological Evaluation, Bureau of Science, Food and Drug Administration, Washington, DC
10. Tyl RW, Marr MC (1996) Chapter 7. Developmental toxicity testing-methodology. In: Hood RD (ed) Handbook of developmental toxicology. CRC Press, Inc., Boca Raton FL
11. Tyl RW (December 4, 2006) Toxicology 707: advanced toxicology regulatory aspects of developmental and reproductive toxicology
12. Guidance for industry considerations for developmental toxicity studies for preventive and therapeutic vaccines for infectious disease indications
13. Pharmaceutical Affairs Bureau, Ministry of Health and Welfare (1984) Japanese guidelines of toxicity studies, notification no. 118. 2. Studies of the Effects of Drugs on Reproduction. Tokyo, Japan: Yakagyo Jiho Co., Ltd.
14. Canada Ministry of Health and Welfare, Health Protection Branch (1973) The testing of chemicals for carcinogenicity, mutagenicity and teratogenicity. The Ministry of Ottawa
15. United Kingdom, Committee on Safety of Medicines (1974) Notes for guidance on reproduction studies. Department of Health and Social Security, London
16. Organization for Economic Cooperation and Development (OECD) (1981) Guideline for testing of chemicals: teratogenicity. Director of Information, Paris
17. ICH Website: History of ICH
18. Food and Drug Administration (FDA) (1994). International Conference on Harmonisation (ICH); Guideline for detection of toxicity to reproduction for medicinal products; Availability; Notice. Section 4.1.3, Study for effects on embryo-fetal development. Fed Register 59(183): 48749
19. Faqi AS (2012) A critical evaluation of developmental and reproductive toxicology in nonhuman primates. Syst Biol Reprod Med 58:23–32
20. Hong Y-S, Kim Y-M, Lee K-E (2012) Methyl mercury exposure and health effects. J Prev Med Public Health 45(6):353–363
21. World Health Organization International (1990) Program on chemical safety. Environmental methylmercury: health criteria 101. Geneva: World Health Organization. Available from: http://www.inchem.org/documents/ehc/ehc/ehc101.htm. Accessed on October 2, 2012
22. ICH harmonised tripartite guideline. Detection of toxicity to reproduction for medicinal products and toxicity to male fertility, S5(R2), 24 June 1993
23. Lerman SA, Hew KW, Stewart J, Stump DG, Wise LD (2009) The nonclinical fertility study design for pharmaceuticals. Birth Defects Res B Dev Reprod Toxicol 86(6):429–436
24. LASA, Guidance on dose level selection for regulatory general toxicology studies for pharmaceuticals
25. CPMP: Note for guidance on repeated dose toxicity. CPMP/SWP/1042/99, July 2000
26. Byers SL, Wiles MV, Dunn SL, Taft RA (2012) Mouse estrous cycle identification tool and images. PLoS One 7(4):e35538
27. Myers M, Britt KL, Wreford NGM, Ebling FJP, Kerr JB (2004) Methods for quantifying follicular numbers within the mouse ovary. Reproduction 127:569
28. Findlay JK, Drummond AE (1996) Control of follicular growth. In: Filicori M, Flamigni C (eds) The ovary: regulation, dysfunction and treatment. Elsevier Science, Amsterdam, pp 13–21

29. McGee EA, Hsueh AJ (2000) Initial and cyclic recruitment of ovarian follicles. Endocr Rev 21:200–214
30. Smith BJ, Plowchalk DR, Sipes IG, Mattison DR (1991) Comparison of random and serial sections in assessment of ovarian toxicity. Reprod Toxicol 5:379–383
31. Tilly JL (2003) Ovarian follicle counts – not as simple as 1, 2, 3. Reprod Biol Endocrinol 1:11–14
32. Regan KS, Cline JM, Creasy D, Davis B, Foley GL, Lanning L, Latendresse JR, Makris S, Morton D, Rehm S, Stebbins K (2005) STP position paper: ovarian follicular counting in the assessment of rodent reproductive toxicity. Toxicol Pathol 33(3):409–412
33. Mangelsdorf I, Buschmann J, Orthen B (2003) Some aspects relating to the evaluation of the effects of chemicals on male fertility. Regul Toxicol Pharmacol 37(3):356–369
34. Seed J, Chapin RE, Clegg ED, Dostal LA, Foote RH, Hurtt ME, Klinefelter GR, Makris SL, Perreault SD, Schrader S, Seyler D, Sprando R, Treinen KA, Veeramachaneni DNR, Wise LD (1996) Methods for assessing sperm motility, morphology, and counts in the rat, rabbit and dog: a consensus report. Reprod Toxicol 10:237–244
35. Bartke A (1999) Role of growth hormone and prolactin in the control of reproduction: what are we learning from transgenic and knock-out animals? Steroids 64(9):598–604
36. Christensen A, Bentley GE, Cabrera R, Ortega HH, Perfito N, Wu TJ, Micevych P (2012) Hormonal regulation of female reproduction. Horm Metab Res 44(8):587–591. doi:10.1055/s-0032-1306301
37. Wilson J (1977) Current status of teratology: general principles and mechanisms derived from animal studies. In: Wilson J, Fraser F (eds) Handbook of teratalogy, 4 vols. Plenum Press, New York, NY, pp 47–74
38. Friedman JM (2010) The principles of teratology: are they still true? Birth Defects Res A Clin Mol Teratol 88:766–768. doi:10.1002/bdra.20697
39. Chellman GJ, Bussiere JL, Makori N, Martin PL, Ooshima Y, Weinbauer GF (2009) Developmental and reproductive toxicology studies in nonhuman primates. Birth Defects Res B Dev Reprod Toxicol 86:446–462
40. Brannen KC, Fenton SE, Hansen DK, Harrouk W, Kim JH, Shuey D (2011) Developmental toxicology—new directions workshop: refining testing strategies and study designs. Birth Defects Res B Dev Reprod Toxicol 92:404–412
41. ICH M3 (R2): Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. Recommended for adoption, June 11 2009.
42. FDA (2006) Guidance for Industry. Considerations for developmental toxicity studies for preventive and therapeutic vaccines for infectious disease indications. FDA, Washington, DC
43. Wise LD, Buschmann J, Feuston MH, Fisher JE, Hew KW, Hoberman AM, Lerman SA, Ooshima Y, Stump DG (2009) Embryo-fetal developmental toxicity study design for pharmaceuticals. Birth Defects Res B Dev Reprod Toxicol 86:418–428
44. Staples RE (1974) Detection of visceral alterations in mammalian fetuses. Teratology 9:37
45. Stuckhardt JL, Poppe SM (1984) Fresh visceral examination of rat and rabbit fetuses used in teratogenicity testing. Teratog Carcinog Mutagen 4:181–188
46. Barrow MV, Taylor WJ (1969) A rapid method for detecting malformations in rat foetuses. J Morphol 127:291–306
47. Wilson JG (1965) Methods for administering agents and detecting malformations in experimental animals. In: Wilson JG, Warkany J (eds) Teratology: principles and techniques. Univ. Chicago Press, Chicago, IL, pp 262–277
48. Kimmel CA, Trammell C (1981) A rapid procedure for routine double staining of cartilage and bone in fetal and adult animals. Biotech Histochem 56(5):271–273
49. Staples RE, Schnell VL (1964) Refinements in rapid clearing technique in the KOH-alizarin red S method for fetal bone. Stain Technol 39:61–63
50. Whitaker J, Dix KM (1979) Double staining technique for rat fetus skeletons in teratological studies. Lab Anim 13:309–310
51. Yasuda M, Yuki T (1996) Color atlas of fetal skeleton of the mouse, rat, and rabbit. Hakushinsha Co., Ltd., Tokyo
52. Young AD, Phipps DE, Astroff B (2000) Large-scale double-staining of rat fetal skeletons using Alizarin Red S and Alcian blue. Teratology 61:273–276
53. Namiki M, Tanaka E, Iwase T, Ikeda Y (1992) Practical technique for double staining of unskinned rat fetuses. Teratology 46:36B
54. Peltzer MA, Schardein JL (1966) A convenient method for processing fetuses for skeletal staining. Stain Technol 41:300–302
55. Tyl RW, Chernoff N, Rogers JM (2007) Altered axial skeletal development. Birth

Defects Res B Dev Reprod Toxicol 80:451–472. doi:10.1002/bdrb.20134

56. Menegola E, Broccia ML, Giavini E (2001) Atlas of rat fetal skeleton double-stained for bone and cartilage. Teratology 64:125–133
57. Makris SL, Solomon HM, Clark R, Shiota K, Barbellion S, Buschmann J, Ema M, Fujiwara M, Grote K, Hazelden KP, Hew KW, Horimoto M, Ooshima Y, Parkinson M, Wise LD (2009) Terminology of developmental abnormalities in common laboratory mammals (version 2). Congenit Anom (Kyoto) 49(3):123–246
58. Martin PL, Breslin W, Rocca M, Wright D, Cavagnaro J (2009) Considerations in assessing the developmental and reproductive toxicity potential of biopharmaceuticals. Birth Defects Res B Dev Reprod Toxicol 86:176–203
59. Teo SK, Denny KH, Stirling DI, Thomas SD, Morseth SL, Hoberman AM (2004) Effects of thalidomide on reproductive function and early embryonic development in male and female New Zealand white rabbits. Birth Defects Res B Dev Reprod Toxicol 71:1–16
60. Bailey GP, Wise LD, Buschmann J, Hurtt M, Fisher JE (2009) Pre- and postnatal developmental toxicity study design for pharmaceuticals. Birth Defects Res B Dev Reprod Toxicol 86:437–445
61. Crofton KM, Foss JA, Hass U, Jensen KF, Levin ED, Parker SP (2008) Undertaking positive control studies as part of developmental neurotoxicity testing: a report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. Neurotoxicol Teratol 30:266–287
62. Clark RL (1999) Endpoints of reproductive system development. In: Daston G, Kimmel C (eds) An evaluation and interpretation of reproductive endpoints for human health risk assessment. International Life Sciences Institute Press, Washington, DC, pp 10–21
63. Morris R (1984) Developments of a water-maze procedure for studying spatial learning in the rat. Neurol Sci Methods 11:47–60
64. Spence SG, Cukierski MA, Manson JM, Robertson RT, Eydelloth RS (1995) Evaluation of the reproductive and developmental toxicity of the AT1-selective angiotensin II receptor antagonist losartan in rats. Teratology 51(6):383–397
65. Spence SG, Allen HL, Cukierski MA, Manson JM, Robertson RT, Eydelloth RS (1995) Defining the susceptible period of developmental toxicity for the AT1-selective angiotensin II receptor antagonist losartan in rats. Teratology 51(6):367–382
66. Waltman R, Tricomi V, Shabanah EH, Arenas R (1973) The effect of anti-inflammatory drugs on parturition parameters in the rat. Prostaglandins 4(1):93–106
67. Holsapple MA, Burns-Naas LA, Hastings KL, Ladics GS, Lavin AL, Makris SL, Yang Y, Luster MI (2005) Proposed testing framework for developmental immunotoxicology (DIT). Toxicol Sci 83(1):18–24
68. Korenbrot CC, Huhtaniemi IT, Weiner RW (1977) Preputial separation as an external sign of pubertal development in the male rat. Biol Reprod 17:298–303
69. Meyer OA, Tilson HA, Byrd WC, Riley MT (1979) A method for the routine assessment of fore- and hindlimb grip strength of rats and mice. Neurobehav Toxicol 1:233–236
70. Adams J, Buelke-Sam J, Kimmel CA, Nelson CJ, Reiter LW, Sobotka TJ, Tilson HA, Nelson BK (1985) Collaborative behavioral teratology study: protocol design and testing procedure. Neurobehav Toxicol Teratol 7:579–586
71. (2011) OECD guideline for the testing of chemicals extended one-generation reproductive toxicity study. Adopted: 28 July 2011. corrected: 2 October 2012 © OECD
72. Tyl RW, Crofton K, Moretto A, Moser V, Sheets JP, Sobotka TJ (2008) Identification and interpretation of developmental neurotoxicity: a report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. Neurotoxicol Teratol 30:349–381
73. Vorhees CV (1997) Methods for detecting long-term CNS dysfunction after prenatal exposure to neurotoxins. Drug Chem Toxicol 20:387–399
74. ICHS3A; note for guidance on toxicokinetics: the assessment of systemic exposure in toxicity studies. S3A.
75. Wier P (2006) Use of toxicokinetics in developmental and reproductive toxicology. In: Hood RD (ed) Handbook of developmental toxicology. CRC Press, New York, NY, pp 571–597
76. Moser VC, McDaniel KM, Philips PM (1991) Rat strain and stock comparisons using a functional observational battery: baseline values and effects of amitraz. Toxicol Appl Pharmacol 108:267–283
77. Chen JJ (2006) Statistical analyses for developmental and reproductive toxicologists. In: Hood RD (ed) Developmental and

reproductive toxicology, a practical approach. CRC, Boca Raton, FL, pp 697–711
78. Faqi AS, Hoberman A, Lewis E, Stump D (2013) Developmental and reproductive toxicology, a comprehensive guide to toxicology in preclinical drug development. Academic, San Diego, CA, pp 335–364
79. Morgan DJ (1997) Drug disposition in mother and foetus. Clin Exp Pharmacol Physiol 24:869–873
80. Van den Bulck K, De Schaepdrijver L, Fransen J, Delille P, Vriens K, Martin P, Kwok D, Powers G, Coussement W (2008) Placental transfer and developmental toxicity testing of a human monoclonal antibody in the rabbit. Reprod Toxicol 26:4
81. Simister NE (2003) Placental transport of immunoglobulin G. Vaccine 21:3365–3369
82. Simister NE, Story CM (1997) Human placental Fc receptors and the transmission of antibodies from mother to fetus. J Reprod Immunol 37:1–23
83. ICH S6(R1) Guideline preclinical safety evaluation of biotechnology-derived pharmaceuticals, 1997.
84. ICH M3(R2): Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals. Draft Step 3 2008. www.ich.org
85. Faustman EM, Gohlke JM, Ponce RA, Lewandowski TA, Seeley MR, Whittaker SG, Griffith WC (2006) Experimental approaches to evaluate mechanisms of developmental toxicity. In: Hood RD (ed) Developmental and reproductive toxicology: a practical approach, 2nd edn. CRC Press, Boca Raton, FL, pp 15–60
86. Karlsson M, Danielsson BR, Nilsson MF (2007) New proposals for testing drugs with IKr-blocking activity to determine their teratogenic potential. Curr Pharm Design 13:2979–2988
87. Coe CL, Lubach GR, Izard KM (1994) Progressive improvement in the transfer of maternal antibody across the order primates. Am J Primatol 32:51–55
88. Pentšuk N, van der Laan JW (2009) An interspecies comparison of placental antibody transfer: new insights into developmental toxicity testing of monoclonal antibodies. Birth Defects Res B Dev Reprod Toxicol 86:328–344
89. Fujimoto K, Terao K, Cho F, Honjo S (1983) The placenta transfer of IgG in the cynomolgus monkey. Jpn J Med Sci Biol 36:171–176
90. Holsapple MP, West LJ, Landreth KS (2003) Species comparison of anatomical and functional immune system development. Birth Defects Res B Dev Reprod Toxicol 68:321–334
91. Van de Perre P (2003) Transfer of antibody via mother's milk. Vaccine 21:3374–3376
92. Ponce R, Abad L, Amaravadi L, Gelzleichter T, Gore E, Green J, Gupta S, Herzyk D, Hurst C, Ivens IA, Kawabata T, Maier C, Mounho B, Rup B, Shankar G, Smith H, Thomas P, Wierda D (2009) Immunogenicity of biologically-derived therapeutics: assessment and interpretation of nonclinical safety studies. Regul Toxicol Pharmacol 54:164–182
93. Van Esch E, Cline M, Buse E, Wood CE, de Rijk EPCT, Weinbauer GF (2008) Summary comparison of female reproductive system in human and the cynomolgus monkey (Macaca fascicularis). Toxicol Pathol 36:171S–172S
94. Buse E, Habermann G, Osterburg I, Korte R, Weinbauer GF (2003) Reproductive/developmental toxicity and immunotoxicity assessment in the nonhuman primate model. Toxicology 185:221–227
95. Weinbauer GF (2002) The non-human primate as a model in developmental and reproductive toxicity. In: Korte R, Weinbauer GF (eds) Primate models in pharmaceutical drug development. Münster, Waxmann Verlag GmbH, pp 49–66
96. Weinbauer GF, Niehoff M, Niehaus M, Srivastav S, Fuchs A, Van Esch E, Cline JM (2008) Physiology and endocrinology of the ovarian cycle in macaques. Toxicol Pathol 36:7S–23S
97. Henck JW, Hilbish KG, Serabian MA, Cavagnaro JA, Hendrickx AG, Agnish ND, Kung AH, Mordenti J (1996) Reproductive toxicity testing of therapeutic biotechnology agents. Teratology 53:185–195
98. Weinbauer GF, Frings W, Fuchs A, Osterburg I (2008) Reproductive/developmental toxicity assessment of biopharmaceuticals in nonhuman primates. In: Cavagnaro JA (ed) Preclinical safety evaluation of biopharmaceuticals. A science-based approach to facilitating clinical trials. John Wiley & Sons Inc., Hoboken, NJ, pp 379–397
99. Hendrie TA, Peterson PE, Short JJ, Tarantal AF, Rothgarn E, Hendrie MI, Hendrickx AG (1996) Frequency of prenatal loss in a macaque breeding colony. Am J Primatol 40:41–53
100. Hendrickx AG, Cukierski MA (1987) Reproductive and developmental toxicology in nonhuman primates. Prog Clin Biol Res 235:73–88
101. Chapman K, Pullen N, Coney L, Dempster M, Andrews L, Bajramovic J, Baldrick P,

Buckley L, Jacobs A, Hale G, Green C, Ragan I, Robinson V (2009) Preclinical development of monoclonal antibodies: considerations for the use of non-human primates. MAbs 1:505–516

102. Buckley LA, Benson K, Davis-Bruno K, Dempster M, Finch GL, Harlow P, Haggerty HG, Hart T, Kinter L, Leighton JK, McNulty J, Roskos L, Saber H, Stauber A, Tabrizi M (2008) Nonclinical aspects of biopharmaceutical development: discussion of case studies at a PhRMA-FDA workshop. Int J Toxicol 27:303–312
103. Buse E (2005) Development of the immune system in the cynomolgus monkey: the appropriate model in human targeted toxicology? J Immunotoxicol 2:211–216
104. Bowman CJ, Breslin WJ, Connor AV, Martin PL, Moffat GJ, Sivaraman L, Tornesi MB, Chivers S (2013) Placental transfer of Fc-containing biopharmaceuticals across species, an industry survey analysis. Birth Defects Res B Dev Reprod Toxicol 98:459–485. doi:10.1002/bdrb.21089
105. Tarantal AF, Hendrickx AG (1988) Characterization of prenatal growth and development in the crab-eating macaque (Macaca fascicularis) by ultrasound. Anat Rec 222:177–184
106. Stewart J (2009) Developmental toxicity testing of monoclonal antibodies: an enhanced pre- and postnatal study design option. Reprod Toxicol 28:220–225
107. Morford LL, Bowman CJ, Blanset DL, Bøgh IB, Chellman GJ, Halpern WG, Weinbauer GF, Coogan TP (2011) Preclinical safety evaluations supporting pediatric drug development with biopharmaceuticals: strategy, challenges, current practices. Birth Defects Res B Dev Reprod Toxicol 92:359–380
108. Piersma AH (2006) Alternative methods for developmental toxicity testing. Basic Clin Pharmacol Toxicol 98:427–431
109. Makris SL, Kim JH, Ellis A, Faber W, Harrouk W, Lewis JM, Paule MG, Seed J, Tassinari M, Tyl R (2011) Current and future needs for developmental toxicity testing. Birth Defects Res B Dev Reprod Toxicol 92:384–394
110. Piersma AH, Genschow E, Verhoef A, Spanjersberg MQI, Brown NA, Brady M, Burns A, Clemann N, Seiler A, Spielmann H (2004) Validation of the postimplantation rat whole embryo culture test in the International ECVAM Validation study on three in vitro embryotoxicity tests. Altern Lab Anim 32:275–307
111. Schmid BP, Trippmache A, Bianchi A (1983) Validation of the whole-embryo culture method for in vitro teratogenicity testing. In: Hayes AW, Schnell RC, Miya TS (eds) Developments and practice in the science of toxicology. Elsevier, Amsterdam, pp 563–566
112. Shin JT, Fishman MC (2002) From zebrafish to human: modular medical models. Annu Rev Genom Hum Genet 3:311–340
113. Spielmann H, Pohl I, Döring B, Liebsch M, Moldenhauer F (1997) The embryonic stem cell test (EST), an in vitro embryotoxicity test using two permanent mouse cell lines: 3T3 fibroblasts and embryonic stem cells. In Vitro Toxicol 10:119–127
114. Chapin R, Augustine-Rauch K, Beyer B, Daston G, Finnell R, Flynn T, Hunter S, Mirkes P, Sue O'Shea K, Piersma A, Sandler D, Vanparys P, Van Maele-Fabry G (2008) State of the art in developmental toxicity screening methods and a way forward: a meeting report addressing embryonic stem cells, whole embryo culture, and zebrafish. Birth Defects Res B Dev Reprod Toxicol 83:446–456
115. Palmer JA, Smith AM, Egnash LA, Conard KR, West PR, Burrier RE, Donley ELR, Kirchner FR (2013) Establishment and assessment of a new human embryonic stem cell-based biomarker assay for developmental toxicity screening. Birth Defects Res B Dev Reprod Toxicol 98:343–363
116. Brown NA, Fabro S (1981) Quantitation of rat embryonic development in vitro: a morphological scoring system. Teratology 24:65–78
117. Klug S, Creech Kraft J, Wildi E, Merker HJ, Persaud TV, Nau H, Neubert D (1989) Influence of 13-cis and all-trans retinoic acid on rat embryonic development in vitro: correlation with isomerisation and drug transfer to the embryo. Arch Toxicol 63:185–192
118. Van Maele-Fabry G, Delhaise F, Picard JJ (1990) Morphogenesis and quantification of the development of post-implantation mouse embryos. Toxicol In Vitro 4:149–156
119. Carney EW, Tornesi B, Keller C, Findlay HA, Nowland WS, Marshall VA, Ozolins TRS (2007) Refinement of a morphological scoring system for postimplantation rabbit conceptuses. Birth Defects Res B Dev Reprod Toxicol 80:213–222
120. Brannen KC, Panzica-Kelly JM, Danberry TL, Augustine-Rauch KA (2010) Development of a zebrafish embryo teratogenicity assay and quantitative prediction model. Birth Defects Res B Dev Reprod Toxicol 89:66–77

Methods in Pharmacology and Toxicology (2017) 117–183
DOI 10.1007/7653_2016_65
© Springer Science+Business Media New York 2016
Published online: 22 June 2016

A Developmental and Reproductive Toxicology Program for Chemical Registration

Kamin J. Johnson, Bethany R. Hannas, Sue Marty, Carol L. Zablotny, Jessica LaRocca, Nicholas Ball, and Amanda K. Andrus

Abstract

The goal of the chapter is to outline the process of testing molecules for potential developmental and reproductive toxicity (DART). Here, the entire process of DART testing is discussed, from the regulatory use of DART data to the conduct and interpretation of the various DART study designs. Although non-animal DART testing strategies are envisioned by the new science of "21st Century Toxicity Testing", these high-content, high-throughput testing paradigms are not sufficiently mature from a scientific perspective to be acceptable on their own by regulatory agencies for chemical registration. Thus, these testing paradigms are not included in this chapter. While the scope of the chapter is broad, it is beyond the range of this chapter to describe all possible scenarios encountered in DART testing. For additional information, the reader is referred to other recent publications on this topic. The chapter is organized in a stepwise manner into three main topics: (1) Use of DART studies for chemical registration; (2) General considerations for all DART study designs; and (3) Descriptions of all DART study designs from a practical perspective, beginning with an initial range-finding study and ending with the complex Extended One-Generation Reproductive Toxicity Study (EOGRTS).

Keywords: DART, Toxicology, Developmental, Reproductive, Chemical, Testing, Extended One-Generation Reproductive Toxicity Study, REACH

1 Use of Developmental and Reproductive Toxicity Studies for Chemical Registration

For industrial chemicals, DART testing is typically motivated by one of three drivers: regulatory compliance; product stewardship; or mode of action (MoA) assessment. Regulatory compliance testing (e.g., chemical registration, specific use legislation such as food contact materials, biocides, and cosmetics) follows a structured approach prescribed by the specific legislation. The information required is typically generated using standard Organisation for Economic Co-operation and Development (OECD) guideline toxicity testing studies performed in accordance with Good Laboratory Practice (GLP). In the majority of cases, the testing approach is tiered, with a base set of information that is supplemented as a consequence of certain triggers being met. For example, the

volume (tonnage) of a substance manufactured or imported is a common trigger for additional information. As manufacturing volume increases, the data requirements increase. In this case the manufacturing volume acts as a proxy for exposure potential. Ideally, all testing should be driven by some exposure metric; unfortunately, there are no existing regulations that base testing needs on an exposure scenario. Instead, exposure proxies such as tonnage [1] or migration into food [2] are used. In addition to this relatively arbitrary trigger, the actual findings from the lower tier studies may trigger the need for further assessment to confirm the presence or absence of effects.

Product stewardship programs within companies and industry associations [3] drive the generation of DART data in a similar manner to regulatory compliance driven testing. These programs typically operate to address chemicals that are not covered by an existing regulatory program, such as chemicals sold into countries where chemical management regulations do not currently exist.

Testing to interrogate the MoA of a potential reproductive or developmental toxicant is typically limited to a small number of chemicals and is not a standard part of other testing programs. The testing is substance or effect specific and utilizes tailor-made study protocols designed to test the MoA hypothesis.

Of the three main drivers, the most significant is regulatory compliance driven testing, and across the different regulatory frameworks, there are many common themes to the DART testing requirements. For industrial chemicals, several regulatory regions have adopted the tiered approach to testing for DART. For example, for a new chemical substance registered under the European Union (EU) Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation, the data requirements for DART are associated with manufacturing or import volume. The minimum data requirement is a reproductive/developmental screening study (OECD 421 or 422), and as manufacturing/import volume increases (or if there are observations in the screening study), developmental toxicity testing in one or two species (OECD 414) and more comprehensive reproductive toxicity evaluations are required (OECD 416 or OECD 443) [1]. In China, a similar tiered approach has been adopted for the registration of new chemicals, but the tonnage trigger for the multigeneration study (manufacturing/import volume) is lower and developmental toxicity testing in two species is not required by default [4].

Another difference between the EU and China new substance regulations is the process that must be followed when generating new data. For a registration in China, one must generate all the data needed and submit these in the registration dossier. However, in the EU there is a desire to reduce animal testing as much as possible while still meeting the goals of providing a comprehensive hazard characterization. Thus when registering a chemical substance in the

EU, the conduct of "higher tier" tests (OECD 414, 416, or 443) can only occur once permission has been granted by the European Chemicals Agency. Consequently, it can be very difficult to comply with the requirements of multiple regulations across multiple geographic regions, and in most situations, one defaults to meeting the needs of the most demanding regulation.

Within the chemical industry, the pesticide sector is currently one of the most heavily regulated in the world, requiring a battery of toxicity tests designed to protect human health. The term pesticide refers to a substance or mixture of substances that is intended for the use of managing, preventing, or repelling pests. Classes of pesticides include, but are not limited to, herbicides for use with weeds, fungicides for control of fungal diseases, and insecticides for the management of insects. The registration of a new pesticide molecule requires a comprehensive assessment of potential toxicity which may need to address global regulatory requirements; similar to industrial chemicals, each country has its own toxicity data requirements for pesticide registration.

In the USA, the Environmental Protection Agency (EPA) is responsible for regulating pesticides under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) [5]. Under FIFRA section 3, initial pesticide registration is specific, and pesticide application is only authorized for use on specific crops/sites with particular application rates and methods. The extensive toxicity data requirements include specific studies to address "hazards to humans, domestic animals, and non-target organisms," and these studies are completed and submitted by the registrant to the EPA for approval. Following approval, pesticides are reviewed for reregistration periodically (approximately every 15 years) and are subject to any new data requirements that may have been established in the interim period.

While different regions may have variations in their risk-assessment approach, in recent years there has been a movement towards harmonization of test guidelines for plant protection products, a sub-class of pesticides. Table 1 compares required testing for several different regions, which demonstrates similar data requirements across the globe.

The output of DART testing, as with all other toxicity testing, is the characterization of potential hazards and the dose level where no hazard is observed in the tested species. This information is then taken forward into two very different applications: classification/labelling and risk assessment. The classification and labelling of hazards has become a cornerstone of chemical management in some regions such as the EU. The main classification and labelling system is the Globally Harmonized System (GHS) for the classification and labelling of chemicals. This was developed as part of a global initiative to bring together and harmonize the different classification and labelling systems employed in various countries

Table 1
Required toxicity tests for plant protection products in different regions

Test	Task guideline	USA	EU	Brazil	Canada	China	India
Two-generation reproductive toxicity—rat	OECD 416	Yes	Yes	Yes	Yes		Yes
Extended one-generation reproductive toxicity (default rat)	OECD 443						Yes (alternative to OECD 416)
Separate male and female studies	OECD 416	Yes	Yes		Yes	Yes	
Dominant lethal assay for male fertility	OECD 478		Yes				
Effects on spermatogenesis	OECD 416	Yes	Yes			Yes	
Effects on oogenesis	OECD 416	Yes	Yes			Yes	
Sperm motility, mobility and morphology	OECD 416	Yes	Yes			Yes	
Investigation of hormonal activity	OECD 416		Yes			Yes	
Teratogenicity test—rat	OECD 414	Yes	Yes	Yes	Yes		Yes
Teratogenicity test—rabbit	OECD 414	Yes	Yes	Yes	Yes		Yes

and geographic regions [6]. In brief, GHS (and other similar systems) lay out a set of criteria for a wide array of toxicological, ecotoxicological, and physical/chemical hazards. Comparing the toxicological data for a chemical to these criteria allows one to assign the relevant hazards to that chemical. Within the GHS, reproductive and developmental toxicity classification is divided into two main categories based on the perceived (or known) relevance of effects to humans (refer to Table 2). Category 1 is assigned to those substances where there is clear evidence of reproductive or developmental toxicity from animal studies (or human data), and it is considered that the observed toxicity would be relevant to humans. This category is further split into 1a (known human reproductive/developmental toxicants) and 1b (presumed human reproductive/developmental toxicants). Category 2 is assigned to those substances where the data are less compelling or there are doubts about the relevance of the observed effects to humans.

Classification of a chemical as a reproductive or developmental toxicant can have significant impacts on how that chemical can be used. This is particularly true in the EU where classification into Category 1 results in that substance being banned from use in certain applications such as consumer products, food contact materials and cosmetics. Considering pesticides, under 1107/2009 in the EU if a pesticide is classified for reproductive of developmental

Table 2
GHS classification criteria for DART

Category	Criteria	Cut-off criteria?
Category 1 Known or presumed human reproductive toxicant	Substances are classified in Category 1 for reproductive toxicity when they are known to have produced an adverse effect on sexual function and fertility, or on development in humans or when 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. The classification of a substance is further distinguished on the basis of whether the evidence for classification is primarily from human data (Category 1a) or from animal data (Category 1b)	Yes/No?
Category 1a Known human reproductive toxicant	The classification of a substance in this Category 1a is largely based on evidence from humans	Yes
Category 1b Presumed human reproductive toxicant	The classification of a substance in this Category 1b is largely based on data from animal studies. Such data shall provide clear evidence of an adverse effect on sexual function and fertility 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 nonspecific 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	Yes
Category 2 Suspected human reproductive toxicant	Substances are classified in Category 2 for reproductive toxicity when there is some evidence from humans or experimental animals, possibly supplemented with other information, of an adverse effect on sexual function and fertility, or on development, and where the evidence is not sufficiently convincing to place the substance in Category 1. If deficiencies in the study make the quality of evidence less convincing, Category 2 could be the more appropriate classification. Such effects shall have been observed 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 nonspecific consequence of the other toxic effects	No

Adapted from ref. [104]

toxicity category 1a or 1b, it cannot be registered. Compare this to the risk-based system in the USA where approval for registration considers both the intrinsic hazard and the potential for human exposure. Under this system, hazard classification alone is not sufficient to prevent registration approval.

Bearing this in mind, perhaps one of the most contentious issues surrounding classification and labelling for reproductive and developmental toxicity is that there is no consideration of potency and risk. The assessment is based solely on whether the observed toxicity is relevant to humans and not secondary to general (i.e., systemic) toxicity. By neglecting to consider potency and risk in the classification and labelling process, many substances that produce effects in animal studies can become excluded from markets where realistically there would be no concern about safe use based upon human exposure levels. Consequently, the interpretation of reproductive and developmental toxicity studies and the assessment of the relevance of any effects to humans is a significant part of the chemical safety assessment. The need to screen early in a toxicity testing program for potential DART effects has become increasingly important in new product development, particularly where the major markets for these substances will utilize a hazard-based approach to chemical management.

1.1 Typical Flow and Integration of Studies for Agrochemicals

The discovery of new candidate pesticide-controlling molecules is critical to mitigate pressures within the agricultural world that are either new or increasing in burden (such as herbicide-resistant weeds). Importantly, pesticides over time have generally both increased their specificity for pest control and decreased their use rates [7]. The general idea of a DART testing program is to identify and characterize potential DART hazards that need to be addressed in the risk assessment process for a molecule.

Not all new candidate molecules make it to market or even through a full testing program, and best practices should be implemented early in the testing program in order to prioritize molecules. To aid in this process, use of recent scientific advances may assist in identifying potential DART hazards. One such predictive tool that may be employed is the DART decision tree [8], which is an in silico tool that uses chemical structure to predict DART activity. Other screening tools that can be used include in vitro estrogen receptor/androgen receptor transactivation assays and the in vitro aromatase assay, which can help identify potential endocrine disrupting compounds. For developmental toxicity prediction, a number of predictive assays have been promulgated over the years including transcriptome profiling of in vitro systems, mammalian whole embryo culture, zebrafish embryo exposures, and embryonic stem cell metabolomics [9, 10]. These predictive tools can be used in a screening program to identify compounds of potential concern, not only just for pesticides but for chemicals in general. A so-called “positive” result in these tools does not necessarily mean that the compound should be removed from the testing program, but rather these data can provide useful information to guide further experimentation in DART testing.

Following the implementation of predictive screens and tools for molecule prioritization, higher tiered testing is necessary to clearly identify developmental and reproductive hazards. Typically, the teratogenicity test in two species (rodent and non-rodent) and two-generation reproductive toxicity in rats are required for registration across the globe. However, prior to conducting these studies a useful screen is the reproductive/developmental toxicity screening test (OECD 421). The OECD 421 is meant to generate "limited information concerning the effects of a test substance on male and female reproductive performance such as gonadal function, mating behavior, conception, development of the conceptus and parturition" (OECD Test Guideline 421). Not only can this study provide data on possible DART effects, but can be used as a dose range-finding study for follow-up testing. Subsequently, the teratogenicity tests can be performed, preferentially in the rat and rabbit (OECD 414). The OECD 414 test guideline is meant to generate data on effects of prenatal exposure in both the pregnant test animal and on the development organism. Additionally, the current dosing paradigm is intended cover both organogenesis and the fetal period, whereas in the past dosing only covered organogenesis. Finally, the two-generation reproductive toxicity study (OECD 416) is a global requirement that is intended to detect adverse effects on the male and female reproductive systems. As the name implies, the two-generation study is meant to assess growth and development of both the F1 generation, as well as the growth, development, and condition and performance of the reproductive systems in the F2 generation. A potential alterative to the two-generation study is the Extended One-Generation Reproductive Toxicity Study (EOGRT) (OECD 443), which is also designed to evaluate potential reproductive and developmental effects following prenatal and postnatal chemical exposure. A major distinction between the EOGRT and the two-generation study is the optional extension of evaluating an F2 generation, which if not included the EOGRT uses significantly fewer animals. However, the two-generation study is the preferred definitive test across most geographies. More detailed descriptions of these test guidelines are given in the "DART Study Designs" section of this chapter.

1.2 Mode of Action Studies

Current approaches for toxicity testing rely on identifying adverse effects that occur at extremely high (relative to human exposure) dose levels in rodents. In recent years, there has been a movement towards a more relevant toxicity testing strategy, such as limiting high dose levels based on toxicokinetic data or relevance to human exposure. However, given today's current toxicity testing paradigm of high dose level testing, adverse effects are occasionally observed in DART studies, such as in the aforementioned OECD 414 or OECD 416. As a general practice, any treatment-related DART effect should be evaluated, often by means of additional

mechanistic data to characterize a MoA. The MoA/Human Relevance Framework (HRF) developed by the International Programme on Chemical Safety of the World Health Organization [11, 12] and the International Life Sciences Institute [13, 14] can be used as a template to define the MoA of a chemical and analyze the human relevance of the MoA. MoA/HRF projects are designed to answer the following general questions: (1) what is the mechanism via which the adverse effects are observed in animals? and (2) are the effects that occur in animals via this mechanism relevant to humans?

1.3 Typical Flow and Integration of Studies for Industrial Chemicals

Reproductive/developmental toxicity assessment for the EU REACH regulation—an example of the workflow of studies and potential considerations when executing the test strategy.

As indicated previously, DART testing in accordance with the requirements of the EU REACH regulation is a tiered process with some studies required and others triggered and then only performed following approval by the European Chemicals Agency (ECHA). In general terms, the data requirements depend on the type of registration required, and that is determined based primarily on the manufacturing or import volume (i.e., tonnage band) for the substance in question. No DART assessment is required for a 1–10 t registration. The requirements for the other levels of registration are given in the Table 3 below.

When preparing a 10–100 t or greater registration dossier, the DART screening study (OECD 421 or 422) should be included in the dossier when it is submitted. For the 100–1000 t and >1000 t registrations, if needed, the "higher tier" studies (OECD 414 and 443) should be proposed and conducted once approved by ECHA. If there are already higher tier studies available for a substance, then the screening study may not be required. Developing and implementing the most appropriate DART testing plan for REACH can be complex, particularly for the higher level registration dossiers. Fortunately, there is an extensive guidance document that leads potential registrants through this process [15] and gives an

Table 3
Industrial chemical DART testing requirements based upon tonnage band

Tonnage band	Reproductive/ developmental toxicity screen (OECD 421/422)	Developmental toxicity (one species—rodent) (OECD 414)	Developmental toxicity (second species—non-rodent) (OECD 414)	EOGRTS (OECD 443)
10–100 t	Yes			
100–1000 t	Yes	Yes	Triggered	Triggered
>1000 t	Yes	Yes	Yes	Yes

overview of the typical considerations that should be discussed when developing any DART testing plan.

One of the challenges with complying with the DART testing requirements for REACH is that the requirements do not include much leeway for minimizing the use of animals. An example of this is the mandatory requirement for developmental toxicity data in two species (rodent and non-rodent). One alternative to this approach would be to perform the reproductive screen and the EOGRTS in rats and the developmental toxicity study in rabbits. Consequently, there will be some form of developmental toxicity data in both rats and rabbits, allowing one to more conclusively characterize the potential hazard. Where there are indications (from the screening study or the EOGRTS) that developmental toxicity is occurring in rats, a second developmental toxicity study could be performed. This is the strategy implemented in the Biocidal Products Regulation in the EU, and it allows for a reduction in the amount of animals needed to assess these endpoints without compromising the hazard assessment.

2 General Considerations for All DART Study Designs

2.1 Principles of the 3Rs (Refinement, Reduction, and Replacement)

The 3Rs refers to Reduction, Refinement and Replacement of animal testing and is a concept that was developed over 50 years ago by Russell and Burch [16] with the aim of improving the treatment of test animals. This framework is widely accepted as a standard to strive for in regulatory chemical testing. In the area of DART, 3Rs are particularly important, as these study types typically involve animal-heavy testing designs, where the litter is the usual statistical unit and therefore uses many animals to obtain a single data point. For chemical registration, roughly 50–70 % of the total animal usage in a testing program for a single compound is attributed to the required guideline DART studies, along with necessary probe or range finding studies [9]. Therefore, a tremendous impact in the area of DART testing would result from any success in: (1) using less animals to arrive at the same conclusions (Reduction); (2) optimizing animal usage to get more information out of the same or less number of animals or improve quality of life (Refinement); or (3) using alternative assays to arrive at the same conclusion (Replacement).

Approaches for alternatives to animal assays are in a period of rapid development and showing some promise for future **replacement** of in vivo studies. Nevertheless, the complexity of reproductive and developmental processes and the current knowledge gaps in comprehensive mechanistic data for DART means there is much to learn before current in vivo guideline studies can be replaced. Therefore, while attention is being directed at development of in vitro and computational batteries, it is prudent to concurrently

evolve existing in vivo DART testing study designs to align with the guiding principles of the 3Rs. Efforts to incorporate these guiding principles into the testing program may include integrating or combining testing to maximize information gained from the minimal number of animals, incorporating endpoints that previously required satellite groups or separate guideline studies, and refining use of test animals to reduce pain, stress, or suffering such as improved blood sampling techniques and selecting the least stressful relevant route of exposure.

As mentioned previously, approaches currently being explored for ultimate **replacement** of animal testing in the field of DART include in silico, toxicogenomic, in vitro and short term in vivo assay approaches. Examples include in vitro nuclear receptor transactivation assays (estrogen and androgen receptor) or small scale in vivo screening assays with targeted endpoints for molecules belonging to a particular class of compounds with known DART activity (e.g., fetal testicular testosterone production, postimplantation loss, or a particular fetal malformation). Currently, many laboratories are exploring and implementing these approaches from a screening standpoint. Use of these screening assays can detect potential DART activity early in the new molecule development stage and thereby help reduce the number of molecules that are only flagged for DART activity by large scale guideline DART studies. From that perspective, these methodologies can serve as replacements for large scale guideline studies on analogues that are not selected for further development based on DART activity detected through screening early in the discovery process.

Reduction in the number of animals used for regulatory DART testing can be accomplished through several different approaches. Laboratories will often order extra time-mated animals for guideline developmental toxicity studies (OECD 414). This practice is carried out to achieve the required number of pregnant animals in each dose group to have an acceptable study by accounting for any non-pregnant animals that may have been mis-identified as pregnant. A retrospective analysis of historical data to assess number of extra animals ordered, compared to the number of non-pregnant animals on studies will potentially allow the laboratory to reduce the number of extra animals that are ordered for each study. Although this seems like a small reduction, over time the number of animals saved is significant when considering both the mother and litter. For example, in reducing group size from 26 time-mated animals/dose group to 24, roughly 72–112 total animals are saved per study when considering litter and dam/doe. Importantly, this reduction in animal numbers does not compromise the study success in any way and, therefore, does not reduce the value added of the remaining study animals. This point is a critical consideration in any sample size reduction. If reduced sample size significantly decreases the sensitivity or success of a study,

the information gained from the animals used is severely diminished. Therefore, a careful cost versus benefit analysis should be conducted prior to any sample size reduction.

To additionally **reduce** animal use in the context of regulatory testing, consideration should be given to the number of animals used in a probe design. In some cases, a triggered design (see below) can optimize information gain versus number of total animals used. For example, when no preexisting toxicity information exists, starting with a single dose level and subsequently triggering additional dose levels up or down based on responses of the initial group can reduce the need to add extra groups after the fact to select an appropriate high dose level [17]. The number of animals per group should also be carefully selected, keeping in mind the cost versus benefit analysis described above.

Another example of a substantial **reduction** in animal numbers is use of the "limit test." The OECD 414 developmental toxicity test guideline recommends a limit test under the following condition: "*If a test at one dose level of at least 1000 mg/kg/body weight/day by oral administration, using the procedures described for this study, produces no observable toxicity and if an effect would not be expected based upon existing data (e.g., from structurally and/or metabolically related compounds), then a full study using three dose levels may not be considered necessary.*" A similar recommendation is cited in the Health Effects Test Guideline of the United States Environmental Protection Agency (OPPTS 870.3700 Prenatal Developmental Toxicity Study). Taken together, both guidelines establish 1000 mg/kg bw/day as the limit dose and provide support for the concept of a "limit test" study, which would utilize controls and 1000 mg/kg bw/day dose groups only. This study design eliminates use of low and mid-dose groups. Use of this approach for the guideline developmental toxicity studies reduces animal numbers by about 48 dams (rat study) and 48 does (rabbit study) plus their respective litters, for a total saving of approximately 1000 animals.

Finally, **reduction** in animal numbers can be achieved by integrating endpoints within a single study design. This concept maximizes the information gained on an animal by animal basis, as opposed to using multiple animals across study types to gain the same information. An example of an integrated study design to reduce animal usage is taking blood, urine and milk samples from probe and guideline study animals for toxicokinetic (TK) analysis, rather than adding satellite animals for this purpose. Using this approach across study types in a testing program will provide important information for understanding the relationship between the systemic dose and toxicity. Additionally, due to innovative sampling techniques, this data can be achieved without the need for satellite animals [18, 19], thus reducing animal demands.

Collection of TK data across study types is an advanced **refinement** in the use of study animals. Specifically, these data can be used to refine study design and help set dose levels in a range that is more relevant for human risk assessment by using kinetically derived maximum dose (KMD) setting (described in more detail below in the Incorporating Toxicokinetics into DART Study Designs section).

Further **refinement** of DART studies includes use of the most human relevant exposure route. Oral gavage is specified in the OECD 414 guidelines and has traditionally been the preferred route for these study types. Nevertheless, gavage bolus dosing typically produces an oscillating daily toxicokinetic profile characterized by a transient high C_{max} followed by low C_{min} values during critical periods of embryo–fetal development. In contrast, dietary administration provides a more constant level of test material intake over a 24-h period based on the natural feeding habits of laboratory animals. This consistent exposure over the dosing period offers more continuous systemic and fetal exposure throughout the study and also more closely represents exposures humans may experience [17, 20]. Additionally, administering test material in the feed spares the animal from stress induced by the daily gavage procedure and eliminates potential animal loss via gavage procedural error.

Comparison studies of exposure routes in the rabbit demonstrated that with a short half-life molecule ($t_{1/2} = 1$ h), there was a striking difference in diurnal fluctuation between C_{max} and C_{min}, with only sixfold difference by dietary exposure, compared to a 368-fold difference by gavage [17, 18]. Additionally, this molecule was detectable only up to 12 h by gavage but up to 24 h by dietary administration. Furthermore, with a longer half-life molecule ($t_{1/2} = 14$ h), diurnal fluctuation was not as significant; however, dietary administration resulted in a twofold higher maximum tolerated dose (nominal) when compared to gavage dosing due to evidence of C_{max}-related maternal toxicity following gavage. Taken together, these data suggest that for compounds amenable to formulating into feed, dietary exposure is a preferable option to gavage administration in DART study designs.

Overall, incorporating 3Rs principles into regulatory safety testing programs for industrial and agricultural compounds improves animal welfare, reduces reliance on animal testing and brings major advances to human exposure relevance of toxicity testing. Therefore, striving towards 3Rs should be a top priority when designing a testing program.

2.2 Species and Strain Selection

The laboratory rat is the most widely used primary species for the assessment of DART activity potential. For developmental toxicity studies, the primary species is the rat, and if required, a second nonrodent species is typically the rabbit. The reproductive toxicity test

guidelines are designed specifically for the rat, and the use of an alternate species requires scientific justification applicable to the specific study in question. Justification for the use of a particular strain applies regardless of the species used, but by way of example, considerations for the rat are described further.

Strain selection is based on a number of factors, with high levels of fecundity and low incidences of spontaneous malformations in the offspring being cited as paramount in the test guidelines. In practice, fecundity is typically of the most concern with common laboratory rat strains. Common rat strains used for toxicology testing are the Fischer, Wistar, and Sprague–Dawley, but caution must be used when comparing strains, as nomenclature and breeding stock origin vary depending on the supplier. The inbred rat strains such as the Fischer rat are less fecund than most outbred strains and are typically not used in DART studies. Of the most common outbred strains, the Sprague–Dawley rat is slightly more prolific than the Wistar rat and demonstrates a higher level of maternal caregiving [21].

Irrespective of fecundity, the two most important considerations for strain selection are the use of a particular strain in previous toxicity studies and the laboratory's experience and availability of historical control data from the selected strain. With regards to previous toxicity studies, data may be available (e.g., acute or subchronic oral toxicity data) which may lead one to select a given strain. This is particularly important when there is a complex toxicity profile for a given substance. It is not uncommon to find differences in toxicokinetics or metabolism between different rat strains, which whether known or unknown can leave the researcher questioning the reproducibility of effects from study to study. How the choice of strain is made may also be dependent on the stage of a testing program. When a testing program is initiated for a new substance, the DART program should be designed around a single strain. This increases the likelihood of the reproducibility of effects from early probe studies to later full studies as well as between developmental toxicity studies and reproductive toxicity studies. Occasionally, a data gap may be identified in a prior set of studies, and as such, the previously used strain should be selected when possible to fulfill the requirements of future studies.

While the previous toxicity information available for a given substance is an important consideration for strain selection, the availability of historical control data (HCD) deserves equal or perhaps greater consideration. The importance of HCD for the interpretation of DART studies cannot be understated, and the data in an HCD set must first be defined by the strain of the rat. This is mandatory given the strong genetic component to many background anomalies in common laboratory species and potential TK differences between strains. With developmental toxicity studies, it is critical to have a robust HCD set given the oftentimes low

incidence of fetal malformations in control rats. Depending on the purpose of a study and the previous toxicity data available, it may be necessary to select a strain or even a testing laboratory with access to a particular strain, based solely on the availability of an adequate HCD set.

Ultimately, the researcher must weigh the risks and benefits associated with the selection of a specific strain depending on the question at hand. It is also important to consider any potential future studies that may be required. Depending on the need for future studies, it may be best to either continue with a previously used strain or switch to a preferred strain early on with substances that require more extensive research.

2.3 Incorporating Toxicokinetics into DART Study Designs

Toxicokinetics (TK) play an integral role in biological activity and potency of a compound. The degree to which a chemical is absorbed into systemic circulation, transformed through metabolism, distributed to various tissue types and eliminated all impact the body's response to exposure. Integration of TK across the comprehensive set of toxicity guideline studies required for registration of agrochemicals and some industrial compounds has lagged behind the pharmaceutical industry, where TK assessments are routine. Nevertheless, including these assessments across study types in the toxicity testing program provides major advantages, as knowledge of systemic exposure levels to the parent compound and its metabolites allows for better understanding of species differences, MoA, and dose response characteristics and for more confident data extrapolation across different routes of exposure [22]. Integrating TK assessments into guideline studies provides a tremendous wealth of information, which strengthens interpretation of data and selection of appropriate dose levels.

Daily internal exposure estimates can be made for dietary subchronic and chronic duration studies based upon area-under-the-curve determinations for plasma or blood concentrations at steady state over 24 h (AUC_{24h}). These AUC_{24h} calculations can be achieved with high accuracy using blood samples obtained from just three time points including C_{max}, mid-point, and C_{min} [19, 23]. Additionally, a rough estimate for elimination half-life ($t_{1/2}$) of the parent and any known metabolites can be determined based on comparing concentrations in a terminal blood sample when animals have been fasted (generally conducted 16–18 h following the removal of the test diet) to blood concentrations at steady state (AUC_{24h} determination). These samples can be obtained from study animals using minimally invasive sampling strategies (e.g., a tail nick) and without the need for additional satellite groups.

For DART study designs, blood samples can be collected from adults prior to breeding, during gestation or at necropsy and collected from fetuses or pups at necropsy. Repeated sampling of dam

blood for AUC_{24h} determination is conducted on the day prior to necropsy (GD 20) in a developmental toxicity probe study and during the prebreeding phase of reproduction studies (OECD 421, 422, 416, 443) for both male and female adult animals. A terminal blood sample is obtained at necropsy via the orbital sinus or vena cava in adults, the umbilical cord in fetuses, or via cardiac puncture in pups. In addition, milk samples for TK analysis can be obtained from dams during the early lactation period on reproduction studies.

TK data serve a variety of purposes in a DART testing program. The data can be used to verify exposure of developing fetuses and pups to the test compound or known metabolites. Comparisons of dam and fetal/pup levels can give an indication about placental or milk transfer of compounds. Additionally, understanding TK thresholds of the tested species should play a key role in dose level setting and interpreting toxicity information. Specifically, the linear range of exposure can be determined based on TK estimates of daily systemic exposure and elimination, such that the point of departure for nonlinearity in internal exposure can indicate saturation of a biological process including absorption, metabolism, biliary elimination, or renal clearance. As supported by OECD Guidance Document No. 116 and the 443 and 416 guidelines, doses above the linear range, indicated by a non-dose-proportional increase in AUC_{24h}, indicates stress on the biological processes of detoxification/activation. Additionally, use of high doses above the linear range can result in effects that are not relevant to the human exposure scenario. Doses above this linear range can induce effects as a consequence of saturating kinetic processes in the high dose range which are unlikely to occur in the lower dose/human-relevant range. For that reason, understanding TK can help refine study design by providing an alternative high dose level setting approach known as kinetically derived maximum dose (KMD) setting, whereby the high dose level is set at or below the point where kinetics become nonlinear [17, 20, 22].

Finally, TK data can be useful in interpreting MoA assessments. Species and/or sex differences or vulnerable exposure windows can be evaluated through TK measurements. Embryonic or fetal exposure can be different from maternal internal exposure due to the degree of placental transfer, as well differences in inherent metabolic capacities. These differences can produce disparities in the toxicity profile between species, sexes, or routes of exposure. Therefore, as it relates to MoA assessment, TK should be one of the first considerations in understanding developmental or reproductive toxicity.

2.4 Dose Level Selection

Selection of appropriate dose levels for DART studies is crucial to obtain useful toxicity data for hazard identification and risk assessment. Dose level selection is outlined below for specific study

designs, but some general approaches are described here. As a molecule progresses through the toxicity testing program, the range of tested dose levels typically becomes lower as more toxicity information is gathered. Dose level selection for DART studies begins with identifying an appropriate high dose level, and lower dose levels naturally follow using appropriate spacing to achieve a dose–response and a No-Observed-Adverse-Effect-Level (NOAEL). Determining an appropriate high dose level is a debated issue with no specific guidance provided by regulatory bodies. DART test guidelines typically state that the high dose level should cause some toxic effects but not death or obvious suffering. Thus, justification of the high dose level is at the discretion of the scientists performing the study.

The high dose level for a DART definitive (i.e., non-probe) study can be selected using the traditional maximum tolerated dose (MTD) approach or using a kinetically derived maximum dose (KMD) approach. The MTD is based on toxic effects such that the high dose produces some systemic toxicity without severe suffering or death (e.g., clinical signs, decreased body weight/gain, or dose-limiting target organ toxicity). In selecting the high dose for a repeat dose study, government, academic, and industry stakeholders indicate that toxicokinetic nonlinearity should be considered [22]. This concept of KMD in selecting the high dose level at or slightly above the inflection point for nonlinear kinetics is supported by OECD test guidelines. In regard to a two-generation reproduction study, the OECD 416 guideline recommends consideration of "*any available information on metabolism and kinetics*" [24]. The OECD 443 guideline for the EOGRTS states that in selecting dose levels "*care should be taken to avoid high dose levels which clearly exhibit saturation, provided of course, that human exposures are expected to be well below the point of saturation. In such cases, the highest dose level should be at, or just slightly above the inflection point for transition to nonlinear TK behavior*" [25]. OECD Guidance Document 116 recommends consideration of kinetic nonlinearity in determination of dose levels: "*Although top dose selection based on identification of inflection points in toxicokinetic nonlinearity may result in study designs that fail to identify traditional target organ or body weight effects, it must be appreciated that metabolic saturation in fact represents an equivalent indicator of biological stress. In this case, the stress is evidenced by appearance of non-linear toxicokinetics rather than appearance of histological damage, adverse changes in clinical chemistry, haematology parameters or decrease in body weight gain*" [26]. Finally, the European Chemicals Agency Guidance for the Implementation of REACH document on repeat dose toxicity study dose level selection states that the dose level at which nonlinear toxicokinetics is observed "*can be regarded as the kinetically derived maximally tolerated dose (MTD)*" [15]. These guidance documents for reproduction toxicity studies and

other repeat dose study designs all reflect the current scientific understanding that dose levels far above the toxicokinetic nonlinear inflection point in animal studies have no relevance to humans when human exposure levels are substantially below the toxicokinetic nonlinear inflection point [22, 27–29].

For the KMD approach to dose level selection, the dose at which TK processes are saturated is determined from prior studies in the species; this point is considered the threshold for transition to nonlinear TK. The high dose is set at, or slightly above, the threshold for nonlinear TK, thereby limiting the opportunity for dose-dependent saturation of TK processes and the subsequent development of irrelevant toxicity data. To use the KMD approach, there must be a sufficient margin of exposure such that human exposures to the compound are well below this nonlinear TK threshold [20, 29]. The method (MTD vs. KMD) for dose selection depends on regulatory requirements and whether an MTD or nonlinear TK occurs at a lower dose level.

3 DART Study Designs

The remainder of this chapter describes the DART testing study designs. These designs are presented in a typical toxicity testing workflow for molecules with no previous repeat dose toxicity data.

3.1 Palatability and Triggered Range-Finding Study in Rats

The first repeat-dose study in a testing program for a new substance is typically the palatability or triggered range-finding study (Fig. 1). While not a regulatory requirement, this non-guideline study is a critical first step prior to the initiation of DART studies. The purpose of this study is to determine the palatability of the test substance for dietary and drinking water routes and to screen for general toxicity. At the conclusion of the study, the data should define a palatable dose level up to a maximum dietary or drinking water concentration that would produce exposure up to the limit dose (1000 mg/kg/day) or identify dose levels (up to the limit dose) producing general

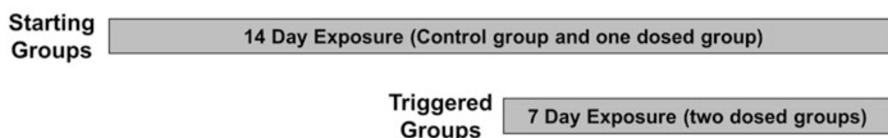

Fig. 1 Experimental design of the palatability and triggered range-finding study. A control group and one dosed group (typically 500 mg/kg/day) are started first, followed by two additional, triggered dose groups 7 days later. All groups are age matched for comparisons of organ weights at study termination on day 14. This design enables determination of an adequate high dose level for subsequent studies with minimal use of animals

toxicity that will assist in high dose level setting for subsequent DART studies.

The design of this study is dependent on the amount and quality of previous toxicity data as well as the purpose of the study. Typically, the study is between 7 and 14 days in duration and is conducted in nonpregnant females only. The choice of females is preferred as they are more relevant to DART studies, and these females should be nonpregnant in order to reduce animal (i.e., fetal) usage. Based on the situation, one of two testing strategies may be used: (1) when there are minimal/no repeat-dose toxicity data, a triggered design is used; or (2) when route-specific data are present but insufficient, a traditional (non-triggered) design can be conducted. Oftentimes, the former is used for new substances (e.g., new pesticide registration) and the latter may be used more for data gaps or REACH registration.

A traditional range-finding study is defined as a study with three or fewer dose groups plus a vehicle control group wherein all groups are initiated at the same time. If more than three dose groups are needed, this suggests that previous data are not sufficient and a triggered design should be used instead. Examples of when a traditional study may be warranted are when ample data are present for a substance, but the data are in a different rat strain or from a different testing laboratory. In other instances, previous toxicity data may be reliable but are outdated, and the researcher must validate that a high dose level is adequate, yet does not exceed an MTD. Under such circumstances, the range finding study is a prudent use of animals, as initiating guideline or probe studies in pregnant animals with larger group sizes is not a casual undertaking.

A triggered range-finding study is defined as a study wherein a single dose group and a vehicle control group are initiated simultaneously, followed by the "triggering" of lower and/or higher dose groups as needed depending on the observed palatability or toxicity (Fig. 1). For molecules with no repeat dose toxicity data, the first dose level chosen is typically 500 mg/kg/day, since it represents one-half of the limit dose. While more complex than a traditional study, the triggered study offers the advantages of reducing the number of animals as well as refining the study such that animals are not unnecessarily over-dosed and dose levels do not need to be terminated prematurely. This is because a traditional (non-triggered) study design in the case of a new substance must always include the limit dose of 1000 mg/kg/day, which may exceed the MTD.

Regardless of the design, a sample size of three to five adult females per group is usually adequate for a palatability or range-finding study. Body weights and feed consumption should be collected daily in all animals to evaluate palatability for dietary or drinking water exposure routes. Both water and feed consumption should be collected if the study is via the drinking water route of

exposure. Data interpretation can be enhanced by collecting pre-treatment body and feed weights, as well as collecting these data in groups that have yet to be placed on study in the case of a triggered range-finding design. Daily handheld clinical observations should also be conducted during the treatment period to closely monitor animal health status. If the study is via oral gavage, then it is best to observe animals both prior to dosing and an hour or more post-dosing. Additional parameters evaluated in the study include a complete gross necropsy with collection of liver and kidney (both organs weighed) and relevant gross lesions for histopathology analyses, as well as any other endpoints that may be appropriate based on previous or analog toxicity. A minimalist approach with regards to the number and type of endpoints is usually best with this study, as data interpretation can be confounded at dose-levels approaching or exceeding an MTD. Often, this study obtains repeat dose toxicity data at the highest dose level used for a test material, and such high dose data may not be relevant for human health risk assessment.

For a triggered range-finding study design, the selection of subsequent dose levels occurs on day 7 of the treatment period and is based solely on the in-life data collected (feed/water consumption, body weight, and animal observations). At times, the first initiated dose level may exceed an MTD and may require termination prior to day 7. Therefore, a diligent study team is required to analyze the data in real time. There are three options for triggering dose levels. When there are no clear effects of treatment at the initial dose level tested (e.g., 500 mg/kg/day), two higher dose levels are selected (Option 1; e.g., 750 and 1000 mg/kg/day). An equivocal or slight effect of treatment at the initial dose level results in the selection of dose levels bracketing this dose level (Option 2; e.g., 250 and 750 mg/kg/day). If the effects of treatment on the first dose level are moderate to severe, the next two doses triggered will be lower than the first (Option 3; e.g., 100 and 250 mg/kg/day).

The data obtained in a triggered range-finding study should be interpreted carefully, as the nature of this study design can make direct comparisons between the groups complex if the study is not conducted appropriately. It is important to have all animals on the study age-matched at the beginning of the treatment period for the first initiated groups in order to make direct comparisons of organ weights at necropsy. Also, given the study is typically conducted in young adult animals to match guideline requirements of future DART studies, the animals are at a steep incline on their growth curve. Therefore, having all groups age matched allows for direct comparisons of body weight and feed consumption between the controls and the triggered groups. While some dose response data can be gleaned from this design, the differences in the treatment duration can make interpretation difficult. The triggered study is

most useful as a comparison of each individual dose group to the control.

Other caveats to this study design include the evaluation of daily in-life data. It is not uncommon for individual animals to gain minimal body weight or even lose a small amount of weight from day to day. The time of day the data are collected should be kept consistent from day to day to avoid fluctuations in the data that are not reflective of treatment. Feed consumption is often a more sensitive end point than body weight, as changes in organ weights (e.g., increased liver weight) can mask decreases in body weight. Animal housing should be monitored carefully each day for signs of the animals scratching feed due to palatability effects. Given the small sample size in this study design and the different durations of the treatment periods among the groups, statistical analyses of these data are not recommended.

At the end of the study, the results should be clear and provide the researcher with information needed to select an appropriate high dose level for later DART probe studies. The triggered range-finding design, while more difficult to execute, has distinct advantages over the traditional range-finding study. It is a refinement to the traditional study in terms of animal welfare. Fewer animals are needed, and there is a decreased likelihood of animal stress or early termination of a dose group due to unnecessarily testing at dose levels exceeding an MTD. It is also beneficial to the researcher, in that it helps narrow the dose range to target for the selection of an acceptable (and palatable) high dose level. A case example would be for a substance where a dose of 750 mg/kg/day is the optimal high dose level for subsequent studies. A triggered range-finding design would provide the necessary preliminary data for high dose level selection from only 12 animals dosed at 0, 500, 750 and 1000 mg/kg/day. If a traditional range-finding study was initiated, the doses may be spread much further apart, say 0, 100, 300 and 1000 mg/kg/day. At the end of the traditional study, the conclusion would be that 1000 was too high and that 300 is a No-Observed-Effect-Level. Additional treatment levels and a control group would need to be tested to identify an acceptable high dose level for later studies, and, furthermore, there would have been no useful information provided by the 100 mg/kg/day group. Based on these significant advantages, the triggered study design should always be considered prior to conducting a traditional study.

3.2 Developmental Toxicity Probe Study

The developmental toxicity probe study (Fig. 2) is a non-guideline study conducted with two major objectives: (1) selection of the high dose for the definitive (i.e., full) developmental toxicity guideline study (OECD 414) using maternal toxicity endpoints and (2) ensuring sufficient numbers of fetuses will be available for examination on the definitive study at the selected high dose level. Keeping these goals in mind, a limited developmental toxicity probe study is

Fig. 2 Experimental design of the rat and rabbit developmental toxicity probe and definitive studies. A control group and three treated groups are dosed beginning just after implantation and continuing until just prior to parturition. The *gray areas* of each bar depict the exposure timing. For studies using the oral gavage exposure route, dams/does are not administered test material on the morning of necropsy (GD 21 in rats and GD 28 in rabbits). For studies using dietary or drinking water exposure routes, animals in treated groups are provided feed or water containing test material until the time of necropsy

completed to ensure that endpoints examined during the definitive study are assessed at appropriate dose levels. While the probe study is conducted in support of the definitive study and is necessary for a properly conducted definitive study, the design of the probe is not driven by a specific guideline. Therefore, several different study designs are utilized across laboratories. In general, the probe study includes a much smaller sample size (five time-mated animals/group) and usually does not collect sufficient information to make strong conclusions regarding developmental toxicity. Use of nonpregnant animals to set dose levels for a developmental toxicity definitive study is not recommended. Pregnancy can alter the response of a number of different endpoints used to determine an appropriate high dose level; therefore, only the pregnant model will serve as a true representation of how the test animal will respond to the test material.

The most relevant exposure route is selected for the definitive study based on expected human exposure, and the probe is conducted via the same route. Consequently, the probe can also provide preliminary information on how well the maternal animals handle the test material in a particular route of exposure. For example, rabbits are known to be particularly sensitive to changes in diet flavor and composition [30]. A dietary developmental toxicity probe will provide an indication about the palatability and acceptability of the test diet to the pregnant doe. Dose levels for the probe are selected with the primary goal of identifying an acceptable high dose for the definitive guideline study. Previous repeat dose toxicity data, preferably in the same species, will guide the selection of doses that are skewed toward the MTD. Dose spacing for the probe is generally closer than the dose selection guidance for the definitive study (1.5–2× spacing for a probe study versus 3–5× spacing for the definitive guideline study), so that the threshold for maternal toxicity is more likely to be identified.

Keeping the two goals mentioned above in mind, a typical probe study design is as follows. Five time-mated animals/group (three treated groups and one control group) are administered test material via the chosen route beginning just after the time of embryonic implantation and ending just prior to parturition (gestation days (GD) 6–21 in the rat and GD 7–28 in the rabbit). Age of the animals at the time of mating follows the attainment of sexual maturity by a few weeks to ensure the period of peak reproductive performance (~10 weeks in the SD rat and ~5 months in the New Zealand White (NZW) rabbit). The assessed in-life parameters include clinical observations, body weights, body weight gains, and feed consumption to examine general maternal toxicity. Observations of the maternal animals are conducted daily and include both a cage side examination and a more detailed hand-held examination designed to detect significant clinical abnormalities and monitor the general health of the animal. Body weights are recorded on GD 0 by the supplier, daily during the dosing period and on GD 21 (rat) or 28 (rabbit) for a terminal body weight. Feed consumption is recorded daily beginning a few days prior to test material administration, through the final day of the study. Toxicokinetic data collection can also be included during the in-life portion of the probe study and/or at necropsy, as these endpoints may offer valuable information with regards to appropriate dose level setting for the definitive study. The postmortem examination includes a gross necropsy and a detailed examination of the reproductive tract (number and position of implantations, viable fetuses and early or late resorptions recorded). Liver and kidney weights are recorded, as these are common target tissues for many industrial and agricultural compounds. For females with one or more viable fetuses, the ovarian corpora lutea are counted and a fetal external examination is conducted on all fetuses. Fetuses are then euthanized and discarded. Visceral, skeletal, craniofacial examinations and fetal weights are not typically included in the probe study design because the emphasis is primarily placed on identifying maternal toxicity. Additionally, the small sample size (five) precludes making definitive conclusions on these fetal endpoints. If no visible implantations are present, pregnancy status is verified by staining the uteri with an sodium sulfide solution to visualize potential early resorptions [31].

Consideration of previous toxicity information and/or physicochemical properties of the test substance can play a critical role in the probe study design, as occasionally this information indicates additional endpoints that should be added to detect maternal toxicity. For example, if the compound is known to be irritating, with previous data indicating point of contact irritation (gross stomach lesions after oral exposure or nasal lesions following inhalation exposure), saving these target tissues for histopathology may be included in the protocol, and these data are used to select dose

levels in the definitive study. Previous toxicity data can also aid in selection of additional endpoints to be added to the study design that may help in mechanistic understanding of maternal toxicity and the relation to fetal outcomes. As an example, moderate anemia induced by hemorrhage was demonstrated to cause increased early resorptions [32]. If the potential exists for the top dose to be limited by anemia or if this endpoint might help interpret developmental toxicity outcomes, then hematology should be included within the developmental toxicity probe study design.

Data analysis and interpretation on the probe study is handled differently than the definitive developmental toxicity study due to the low numbers of animals included per group on the probe study. The small group sizes in the probe study provide low statistical power increasing the likelihood of identifying false positive or false negative effects. Thus, only descriptive statistics (mean ± standard deviations) are collected on all and these data are interpreted with caution. Determination of treatment-related effects is based on sound scientific principles such as biological plausibility, comparison to concurrent and historical control data, and dose–response relationships. Equivocal findings are noted and addressed in the definitive developmental toxicity study, which includes much higher statistical power due to the larger group sizes.

3.3 Developmental Toxicity Definitive Study (OECD 414)

The purpose of the definitive developmental toxicity guideline study (Fig. 2) is to evaluate maternal and developmental toxicity of the test material. Although this study design is driven primarily by the relevant test guidelines (e.g., OECD 414), different labs have slight variations in the conduct of the study. Nevertheless, in all cases detecting toxicity to the developing fetus in the form of death, structural malformations, or altered growth is the principal objective and is carried out through a detailed examination of each fetus.

Dose selection should be based upon previous toxicity information provided by a developmental toxicity probe conducted under similar conditions (same species, strain, route of exposure, vehicle, and laboratory), as described in the preceding section. Using these data, a high dose level is selected that will show maternal toxicity but not induce severe maternal suffering or death. Typically, maternal body weight data are used to select a high dose level that will cause no more than a 10–20 % decrease in body weight gain during 1 or more 3 day measurement intervals. Interpretation of probe study maternal body weight decreases during the last week of gestation in the rat must be done with caution since a decrease may be caused by fetal (not maternal) body weight reduction and/or decreased fetal numbers. Evaluating maternal body weight data in conjunction with feed consumption data may aid in distinguishing maternal and fetal effects. The high dose can also be driven by palatability of the test material on a dietary study, whereby a dose level above the threshold of palatability will reduce

feed consumption and therefore drive maternal body weight and body weight gain down. In rabbits, this issue is somewhat mitigated by addition of “Generally Recognized As Safe”-certified apple flavoring to the test diet to improve palatability. Another factor to consider in the dose level selection is TK data. Specifically, TK data from probe studies can be obtained to identify levels of parent compound or metabolites reaching the fetus. These data can be important for both dose level setting (via the KMD approach) and interpretation of findings. A goal of the developmental toxicity study is to establish the maternal and fetal NOAEL, and therefore, this goal should be kept in mind when selecting the low dose. In certain instances, it may be necessary to repeat a particular dose level from the probe study to determine the relationship to treatment of equivocal developmental effects observed in the probe study. The low number of litters assessed during the probe may make it difficult to interpret findings within the context of the study and therefore, additional information at the same dose level on the full study can aid in distinguishing treatment-related versus incidental effects. For example, observation of a rare external fetal malformation in a single high dose litter on the probe may warrant repeating that dose level on the definitive study to help interpret the finding as either treatment-related or spurious and unrelated to treatment.

The basic study design for the definitive study is similar to the above-described probe study in terms of dosing duration (GD 6–21 for rats and GD 7–28 for rabbits), in-life parameters and certain aspects of the postmortem examination. However, the definitive study includes more litters per group and more endpoints, including an extensive fetal examination. Twenty-four time-mated animals/dose group are included in this study design. This number is selected to target approximately 20 pregnant animals (or animals with implantation sites) at necropsy. Daily hand-held clinical observations are conducted and body weight and feed consumption data are recorded as in the probe study. Postmortem examination of all dams/does is performed as in the probe studies. The sequence of maternal necropsies is counterbalanced across groups (e.g., control, high) to control for potential confounding influences of timing on skeletal ossification [33, 34]. If TK is included in the study design, a terminal blood sample is taken from the first four pregnant females/group and their respective litters at necropsy on GD 21 (rat) or 28 (rabbit).

Fetal examinations consist of the sex and body weight determination and an external examination on all fetuses. Additionally, visceral, skeletal, and craniofacial exams are performed on all (rabbits) or a roughly half (rats) the fetuses per litter. The most commonly cited visceral examination techniques include that of Staples [35] and Wilson [36] which are specified in the test guidelines. In brief, the Wilson technique involves fixation of the fetus in

Bouins solution at the time of necropsy and subsequent gross exam by freehand serial sectioning with a scalpel through the fixed fetus trunk and head regions. The Staples technique involves an examination of the fresh fetus under a dissecting microscope. This technique includes a methodical evaluation of major visceral organs and features, and a detailed assessment of the major vessels and internal structures of the heart. Many sources have described the advantages and disadvantages to each technique [37, 38]. All rabbit fetuses and roughly half of the rat fetuses in a litter are thoroughly examined for visceral abnormalities according to methods based on Staples [35] and Stuckhardt and Poppe [39]. The heads of roughly half of the rat and rabbit fetuses in each litter are removed, placed in Bouin's fixative and serially sectioned to allow for inspection of the eyes, brain, nasal passages, and tongue according to the technique of Wilson [36].

All rabbit fetuses and the remaining rat fetuses not selected for visceral examination are prepared, stained and examined for skeletal abnormalities according to methods based on Trueman [40] and Zablotny [41] for double staining or Dawson [42] for single staining. In brief, fetuses are eviscerated and skinned to allow penetration of the stain(s). Next, they are fixed in ethanol, then macerated, stained and cleared. The single staining technique utilizes Alizarin Red-S and stains only ossified bone, leaving it a red-purple color. Double staining targets ossified bone with Alizarin Red-S and cartilaginous tissue, which is stained using Alcian Blue. The fetal staining process can be accomplished using a manual method in which solutions are poured into and drained from compartmentalized boxes. Alternatively, use of an automated pathology tissue processor greatly increases the efficiency of the maceration, staining and clearing of eviscerated and skinned fetuses [41]. The skeletal examinations are performed by careful evaluation of all bones from head to the tail, from both the ventral and dorsal positions. Assessments include presence or absence of a bone, degree of bone ossification and any abnormality in bone structure including misshapen bones, misalignment or lack of symmetry. Double-stained specimens have the advantage of showing underlying cartilage, which can help distinguish between an under-ossified and a missing bone.

Proficiency and skill of the evaluators is paramount for visceral and skeletal examinations, as familiarity of normally developed structures will allow for recognition of altered development. Identification of sometimes extremely subtle differences in structure is necessary, as in the case of degree of bone ossification. Extensive visceral and skeletal training, including assessment of proficiency by way of written and practical exams ensures all examiners are well equipped to identify any alterations in the fetus. All fetal alterations are classified as either a variation or malformation. A variation is defined as divergence beyond the normal range of structural

constitution that may not adversely affect survival or health, whereas a malformation is defined as a permanent structural change that may adversely affect survival, development, or function [43]. Certain variants are seen with relatively high frequency among control animals and may therefore be defined within the laboratory's Standard Operating Procedure as within normal limits and not recorded during examinations. For example, skeletal variants routinely observed in control animals include: pinpoint foramina of skull bones and lumbar or cervical vertebral spurs in rats and rabbits, holes in the xyphoid process of rats, and bone islands in the skull or a single band of cartilage between the fifth and sixth sternabrae in rabbits.

The different stages of embryonic or fetal development can have varying degrees of susceptibility to certain compounds. Therefore, treatment-related resorptions or fetal deaths are important indicators of disrupted development and can represent maternal toxicity-mediated effects or direct embryo–fetal toxicity. Postimplantation loss can represent a complete picture of embryo to fetal toxicity that may occur during the dosing period, as this endpoint accounts for early and late resorptions as well as dead fetuses. Early resorptions can occur in response to embryo toxicity of the test compound resulting in death and consequent resorbing of the embryo. Toxic insult which progresses beyond the early stages of gestation but ultimately results in resorption of the fetus may leave fetal and placental tissue in the uterus. This finding is classed as late resorption at the time of necropsy. Finally, while resistant to induction of major lethal malformations earlier in gestation, a fetus may succumb to fetal toxicity just prior to the end of gestation and be classed as a dead fetus. Viable fetuses with treatment-related structural abnormalities are another indicator of teratogenic properties of a compound; keeping in mind that severely malformed fetuses may be resorbed during gestation and therefore, postimplantation loss and malformations must be interpreted together.

In addition to embryonic/fetal demise and malformations, developmental delay can occur, most often as an accompaniment to maternal toxicity. One key tenant of interpreting developmental toxicity data is dependence on the Weight of Evidence (WoE) approach. Multiple endpoints are assessed and must be viewed collectively in order to identify patterns and overall plausibility of the induced effects. Fetal weights are always interpreted in light of litter size, as the two parameters are generally inversely related. In addition, evaluation of the corrected maternal body weight (maternal body weight minus gravid uterine weight) can give an indication about the potential secondary nature of fetal weight decreases to maternal toxicity. If maternal weights minus the gravid uterine weights are similar between treated and concurrent control groups, this would suggest that decreased fetal weights in a treated group is independent of maternal toxicity and likely driven by direct fetal

toxicity. In other words, smaller fetal size would lead to a smaller gravid uterine weight which would drive the maternal weight lower. Therefore, correcting for the gravid uterine weight would distinguish true effects on maternal weight from effects on gravid uterine weight (due to contents of smaller fetuses). A consideration of sex ratio can play another role in fetal weight means, as male fetuses are slightly larger than females in both rats and rabbits. Therefore, careful consideration of the weights by sex as well as sexes combined may offer further important information for interpreting fetal growth.

With regards to the interpretation of fetal developmental toxicity, one of the most important factors from a risk assessment and classification and labelling perspective is the relationship of fetal effects to maternal toxicity. Adverse developmental outcomes that occur at doses that do not elicit maternal toxicity are of greater concern, as this signifies a selective toxicity on the developing conceptus [44–46]. On the contrary, developmental effects on the offspring that occur at dose levels where the mother is also affected can signify fetal toxicity that is induced secondary to maternal effects. The embryo/fetus is dependent on the maternal compartment for an array of physiological needs, and therefore, effects on the mother can intimately affect in utero development. Nevertheless, without definitive mechanistic information, the causal link between maternal toxicity and fetal development can oftentimes only be hypothesized.

One area in which extensive research has been conducted to evaluate outcomes on the fetus is maternal feed consumption (and associated body weight decrements) and subsequent fetal development. Numerous maternal feed restriction studies have demonstrated the interrelatedness of maternal feed intake and reduced fetal body weight and/or delays in bone ossification (i.e., fetal developmental delay). Data from published feed restriction research and internal historical data can be synthesized into a few major rules to aid in data interpretation of developmental toxicity study data. Specifically, synthesis of the rat feed restriction model data reveals several conclusions regarding the relationship between maternal body weight gain and feed consumption decrements and fetal toxicity:

- Decreased gestation day (GD) 6–21 feed consumption up to 50 % and decreased maternal body weight gain from GD 7–21 up to 60 % have no treatment-related effect on visceral exam endpoints, external exam endpoints, litter size, gestation length, sex ratio, gestation survival, or postimplantation loss [47, 48].
- Fetal body weight at GD 21 is relatively insensitive to decreases in maternal feed consumption or maternal weight gain during GD 6–14. A 50 % decrease in feed consumption and <10 % body weight loss during the first week of an OECD 414 test

(GD 6–14) has no discernible effect on embryo/fetal development [47, 49].

- GD 14–21 (the final week of gestation when fetal body weight increases exponentially) is the sensitive window for fetal body weight decreases in relationship to decreases in maternal feed consumption and body weight gain. A decrease in feed consumption of more than 30 % but less than 50 % during the last week of gestation (GD 14–21) produces a significant reduction in fetal body weight of 15 % [49]. A decrease in maternal body weight gain of more than 15 % but less than 35 % during the last week of gestation (GD 14–21) produces a significant reduction in fetal body weight [47, 48]. Zhang et al. [50] showed that a 50 % decrease in feed consumption accompanied by a 22 % decrease in maternal body weight gain during GD 14–21 significantly reduced postnatal day one pup body weight by approximately 15 %.
- Delays in bone ossification are less sensitive than fetal body weight decrements to decreases in feed consumption or maternal body weight gain. Decreases in feed consumption and maternal body weight gain during the final week of gestation ≥70 % produce a delay in fetal bone ossification [48].

Interpretation of developmental delays or alterations in growth of the fetus should be considered in the context of maternal toxicity, pattern of effects, and degree of delay (i.e., minor skeletal variations versus missing bone or cartilage anlagen). Fetal skeletal variations typically observed are delays in bone ossification and are often associated with delay in maternal body weight gain [51, 52]. These effects combined represent a generalized delay in fetal maturation (i.e., developmental delay). In rats, the days just prior to parturition are the peak periods for rapid bone ossification [33, 34, 51], especially in the phalanges, thoracic and lumbar vertebral centra, sternebrae, and calvarium. Therefore, a pattern of incomplete or delayed ossification limited to these late-ossifying bones is more likely associated with a generalized developmental delay than with skeletal dysplasia [51, 53]. Multiple studies have indicated that delayed ossification is transient, and fetuses with delayed bone ossification will undergo normal bone ossification after parturition [51, 54–56].

3.4 Reproduction/Developmental Toxicity Screening Study (OECD 421/422)

The reproduction/developmental toxicity screening test (OECD 421 or 422) (Fig. 3) is performed as a requirement under European REACH legislation for industrial chemicals with an annual production or import level of 10 t or more. Conversely, these studies are not a requirement for pesticide chemical registration; instead, the definitive multigeneration (OECD 416) or the EOGRTS (OECD 443) is the requirement. Often, the reproduction/

Pre-Exposure	Pre-Breeding	Mating	Gestation	Lactation

Fig. 3 Experimental design of the rat reproduction/developmental toxicity screening study. A control group and three treated groups are administered test material continuously prior to breeding, throughout mating and gestation, and up to approximately lactation day 14. Parental males are administered test material for a total of at least 28 days

developmental toxicity screening study is performed during the pesticide toxicity testing program to identify an appropriate high dose for the required definitive reproduction study (two-generation; OECD 416) and to obtain TK and endocrine activity information earlier in the testing program.

These screening studies were explicitly designed to serve solely as an indicator of potential for reproductive and/or developmental toxicity and were not intended to be used in place of the definitive study designs (OECD 414, 416, and 443). The study design results in data that is useful in priority setting for further definitive testing. The study design is such that for both the OECD 421 and 422, a minimum of eight pregnant rats and ten males are exposed beginning at least 2 weeks prior to mating, through mating (up to 2 weeks), and through at least 28 days total exposure duration for males or through gestation, early lactation, and until necropsy (lactation day (LD) 14) for females. The study designs cover gamete release, fertilization, early preimplantation embryogenesis, implantation, full embryogenesis, fetal development, parturition, and postnatal development up to LD 13 (ending during the period of peak milk production by the dam). The OECD 421 study specifically focuses on reproduction and development, whereas the OECD 422 design is a combined screening test that covers all the endpoints of the OECD 407 Repeated Dose 28-day Oral Toxicity Study in addition to the endpoints included in the OECD 421 study.

The OECD 421/422 studies are terminated after LD 13, and therefore, effects on sexual maturation or reproductive capacity of the offspring cannot be completely determined. Gamete production in males is not fully assessed in the screening studies, as the exposure period does not cover an entire spermatogenic cycle. Additionally, teratogenicity cannot be definitively ruled out, as visceral and skeletal examinations are not performed on the offspring, and maternal cannibalism of malformed offspring during the postnatal period may go undetected. Additionally, a smaller sample size is used in the screening studies than is included on the definitive reproductive or developmental toxicity guideline studies (OECD 443, 416 or 414) and results in lower statistical power for the majority of reproductive endpoints. With these factors taken together, the interpretation, relevance and utility of the screening studies must be carefully considered in regards to regulatory utility.

Dose level setting for the OECD 421/422 studies should be based upon data from previous studies conducted in the same rat strain. As in most guideline study types, the aim for setting the high dose in the OECD 421/422 is to achieve an acceptable high dose level through induction of "toxic effects but not death nor obvious suffering." The guideline also suggests consideration of any available TK data in dose level setting, geared toward using a KMD. Typically, TK data generated within the context of the OECD 421/422 is considered in using the KMD approach for the definitive study and in this case the dose levels in the screening study will be selected based on the criteria for reaching a maximum acceptable dose. For agrochemicals, data generated on the preceding developmental toxicity probe study typically is used to set the reproduction/developmental screening study dose levels. In cases where no suitable preceding data are available, a triggered range-finding study may be necessary. Because this range-finding study is typically performed in nonpregnant animals, the potential for differences in sensitivity of nonpregnant, pregnant, and lactating females to systemic toxicity should be considered in the dose level selection for the OECD 421/422 [57].

Endpoints addressing both systemic toxicity and reproductive/developmental toxicity are assessed under the OECD 421/422 test guidelines. General assessment of the health of the parental generation males and females is included in the OECD 421 study, while a more comprehensive evaluation of general toxicity is covered in the OECD 422 study to achieve the combination of the OECD 407 (28-day repeated dose toxicity test) with the OECD 421 screen. More specifically, systemic toxicity is assessed during the OECD 421 study by monitoring male and female clinical signs through daily observations, body weight, and feed consumption. At necropsy, reproductive and endocrine-specific tissues in adults are preserved for histopathology, along with any tissues showing macroscopic lesions. Offspring are observed daily for signs of toxicity and body weights are recorded. The OECD 422 includes these general toxicity endpoints, along with assessments of hematology, clinical chemistry, functional observations (neurobehavioral assessments) and full histopathology in all adults. Overall, systemic toxicity assessed in the context of the OECD 422 study has much higher statistical power than the 28-day oral toxicity study (OECD 407), as the OECD 422 sample size is roughly doubled in comparison.

Specific DART endpoints that are assessed within the context the OECD 421/422 screening studies include: mating, fertility and conception success, maternal toxicity through gestation and lactation and success of gestation, survival, and growth of the offspring. Male reproductive tract tissues are weighed at necropsy and histopathology is performed on reproductive and endocrine tissues from both male and female adults. In males, a detailed

histological examination of the testes is conducted, which is sufficient for detecting the majority of effects on male fertility and spermatogenesis, following a minimum of 4 weeks of exposure [57–59]. Female fertility and gonadal function is accounted for by measurement of ovary, pituitary and uterine weights, histopathology of these organs plus the vagina, and evaluation of estrous cyclicity. Success of gestation, survival, and offspring growth and development is assessed through postimplantation loss, live litter size and pup survival evaluations, external examinations and pup in-life observations, sex ratio, and pup body weights.

Recently, enhancements have been made to the OECD 421/422 guidelines to increase the ability to detect compounds that are active in the estrogen, androgen, thyroid, and steroidogenesis endocrine pathways. The added endocrine disruptor-relevant endpoints include estrous cycle evaluations of the females prior to conception, thyroid assessments on adults and offspring, and assessment of markers of antiandrogenicity in male offspring including anogenital distance, nipple retention, and visual examination of the external reproductive organs.

In the adult females, vaginal smears are collected for daily evaluation for a minimum of 2 weeks prior to mating and through the mating period until evidence of mating is attained. Additionally, only females that demonstrate normal 4–5 day estrous cycles are included on the study, so pre-exposure estrous cycling is evaluated prior to placing the animals on study. Another complexity to accommodating evaluation and interpretation of this endpoint is that in the rodent, estrous cycling can be affected nonspecifically by acute toxicity factors including reduced feed consumption and body weight [60–63]. Therefore, to acclimate the animals to test material exposure, test material administration is initiated 2 weeks prior to the collection of daily vaginal smears during the 2 week pre-mating period. In all, following arrival of the animals and acclimation to the laboratory for 1 week, females are monitored daily for estrous stage for 2 weeks and then placed on study (at approximately 10 weeks of age) if normal cycling is demonstrated over the 2 week period. Next, exposure to the test compound is initiated for both males and females. During this exposure acclimation period, no smears are performed. Then, females are smeared daily during the 2 week pre-mating period. Daily vaginal smears are performed into the mating period (which can last up to 2 weeks) until evidence of mating is attained (presence of sperm in the vaginal canal or presence of a copulatory plug). Finally, stage of estrous at necropsy (postnatal day (PND) 14) will be determined for each female via a vaginal smear on the morning of necropsy.

Thyroid hormone assessments are conducted at a minimum on all males and two offspring from each litter on PND 13. Initially, T4 measurements are conducted. If changes are seen, additional hormone measurements, thyroid weights and histopathology can be

performed. More specifically, if changes are detected in male and/or PND 13 offspring T4, than T4 should also be measured in dams and blood from PND 4 pups. Next, TSH and T3 can be assessed, as well as measuring thyroid weights (from fixed thyroids) and conducting histopathology on the tissues. A PND 13 assessment of thyroid hormone is a new endpoint which has not been included in previous versions of these guidelines or other testing guidelines. As a consequence, several factors will need to be considered when interpreting the data. Initially, there will be lack of substantial historical control for these measurements. There are age-related changes in thyroid hormone levels in rat pups during lactation. Specifically, T4 levels are very low on PND 4, increase ~13× by PND 15, then decrease ~4× by PND 21 to levels similar to adult animals. T4 levels in normal rat pups are in the range of 0.5–1.0 μg/dl on PND 4 [64, 65], rising to 8–12 μg/dl on PND 15, then declining to adult levels of approximately 3 μg/dl by PND 21 [66]. Developmental delays (e.g., decreased pup body weights) could affect T4 levels by altering the transition to adult T4 levels.

Other endocrine disruptor indicator endpoints that have been adopted into the OECD 421/422 test guidelines include evaluation of offspring for evidence of disrupted androgen-dependent development. These evaluations include measurement of anogenital distance (AGD) in neonatal rats, assessment of male nipple retention and close examination of the external genitalia for abnormalities on PND 13. AGD reduction and nipple retention in male rats have been widely demonstrated as predictive of adverse effects on reproductive tract development, including decreased testosterone production and consequent malformation of the external genitalia and decreased reproductive organ weights [67–70].

AGD measurements are required in the OECD 443 and are triggered in the OECD 416 in the F2 generation only if a treatment-related effect on sex ratio or age at pubertal onset is detected in the F1 generation. To reduce interobserver variability, all efforts must be made to have a single observer perform all measurements within the context of a single study or that measurement techniques are uniform between observers. AGD is influenced by size of the animal, and therefore, it is best evaluated with the cube root of the body weight (adjusted from a three dimensional parameter to be comparable with the one-dimensional parameter of AGD) used as a covariate. In this way, if there is a statistically significant change in AGD, it can be considered treatment-related if it cannot be explained by the size of the animal. From a logistical perspective, this requires pup body weight be collected at the same time as AGD measurements are performed and that these assessments are done on the same postnatal day for all pups (between PND 0 and 4) and ideally by the same observer across an entire study.

In much the same way, evaluation of nipple retention can also present the need for careful coordination from a logistical perspective. Normally, during in utero male sexual differentiation, the areola and nipple formation is reversed through the action of 5α-dihydrotestosterone derived from fetal testicular androgen production [71–74]. Compounds that interfere with in utero androgen production or action will result in retained areola and/or nipples in male offspring. This endpoint is generally assessed on PND 13 pups, as nipples are prominent at this age, and it is also just prior to the thickening of hair growth that could block visualization of any retained structures. Nevertheless, assessment of nipple retention in PND 13 pups is quite subjective. Areolae or nipple retention in the presence of prenatal androgen hormone disruption can display a spectrum of appearance from very faint to prominent and easily identified [75]. In addition, dark focal areas at the nipple anlagen can occur in the absence of any external mammary anlagen (areola and nipple) retention, which can result in an artificially high control incidence of retained nipple counts in an untrained observer. Therefore, it is imperative that staff be well trained to identify retained nipples or areola and ideally, that the same observer evaluates all offspring within the context of a single study.

Hypospadias is a malformation of the urethra in males, where the urinary opening is shifted from the end of the phallus. This malformation is the most commonly cited malformation resulting from exposure to antiandrogen activity. Therefore, careful assessment of the external genitals in PND 13 offspring has been amended to the OECD 421/422 test guidelines. In male rats, separation of the prepuce from the glans penis marks pubertal onset in the male beginning at PND 35–46 and therefore, only a genital bud is present at PND 13. In conducting the external examination of offspring, observers should therefore be aware of the potential for urethral displacement to the base of the genital bud as an early indicator of hypospadias [69, 73, 76].

Due to the subjective nature or relative lack of historical control for many of these endpoints, extensive training and validation within a laboratory may be required to ensure uniformity in collection of the data. Depending on the experience of the laboratory and frequency of conducting these endocrine disruptor-relevant assessments, a positive control study utilizing a strong antiandrogen (flutamide) is likely the most effective method for comprehensive training on recognizing and evaluating these endpoints.

As mentioned above, integration of sampling for TK of the parent test compound and major metabolites can be highly informative and useful when performed in the context of the OECD 421/422 studies. TK data can be collected on male and female adults in the prebreeding period, female blood and milk in lactation, and offspring at termination. Beyond using this data for dose level setting according to the KMD approach, this TK information

is critical to the design of the EOGRTS. The OECD 443 for the EOGRTS states: "If gavage studies are performed, it should be noted that the pups will normally only receive test substance indirectly through the milk, until direct dosing commences for them at weaning. In diet or drinking water studies, the pups will additionally receive test substance directly when they commence eating for themselves during the last week of the lactation period. Modifications to the study design should be considered when excretion of the test substance in milk is poor and where there is lack of evidence for a continuous exposure of the offspring. In these cases, direct dosing of pups during the lactation period should be considered based on available TK information, offspring toxicity or changes in bio-markers." Therefore, milk and nursing offspring TK information obtained during the OECD 421/422 can determine whether the study design modification of direct dosing of the nursing pups is necessary on the subsequent EOGRTS.

Interpretation of reproductive performance should be accomplished using a collective WoE approach, based on the indices of reproductive performance along with organ weight, histopathology and estrous cycling information. Here, the significance of the various indices of reproductive performances is summarized briefly. The mating index can be impacted by neurobehavioral factors, along with endocrine disruptions and therefore, the information collected in the OECD 421/422 is generally not sufficient to determine the cause of an effect on mating. Although histopathology may demonstrate an effect on spermatogenesis in the male testes, an effect on the fertility index may not be reflected, since males are treated for a shorter period than the full spermatogenic cycle and the large excess of spermatozoa in the male rodent requires that a large reduction in sperm number occur before fertility is affected. The gestation index gives an indication of the ability to produce a live litter, but is calculated based on the number of females with at least one live born pup per number of pregnant females. Therefore, this index is not very sensitive as it does not give any weight to number of resorptions in a litter. Reduced gestation length is generally directly related to birth weights and pup survival in that a significant reduction in gestation time leads to reduced pup weights and survival. Conversely, significantly increased gestation length can lead to dystocia, thus adversely affecting pup and dam health and/or survival. Litter size is a fairly sensitive indicator of disrupted reproductive performance. This parameter can be altered by decreased ovulation, disrupted sperm parameters or embryo/fetal demise. Altered sex ratio can indicate selective genetic changes, endocrine alterations or selective demise of one sex. Finally, the pup survival index is determined several times during lactation and provides an indication about ability of offspring to survive postnatally. The survival index can be affected by both direct and indirect effects on the pup. Direct effects include

developmental effects in pups or structural abnormalities and indirect effects can occur through maternal neglect or lack of milk. The survival index may also be affected by litter size (a dam may not be able to fully support an extra-large litter) or birth weight, where smaller pups are more susceptible to reduced survival.

To a certain extent, effects on reproductive performance can be attributed to specific tissue insults in males or females based on the reproductive organ weight and/or histopathology data collected during the OECD 421/422 studies. In males, testicular weights are highly stable (relatively insensitive to body weight changes) and have low variability between individuals. Therefore a change in absolute testis weight generally indicates an adverse effect, which can be further defined by histopathology. The weights of the prostate, seminal vesicles, levator ani plus bulbocavernosus (LABC) muscle complex, and Cowper's glands are all androgen-dependent. Therefore, altered androgen-dependent tissue weights can indicate disrupted testicular function or endocrine impairments. In females, the ovary weights and histopathology can reveal many different adverse effects including follicular depletion, oocyte toxicity, disrupted ovulation, and altered corpus luteum formation, changes in folliculogenesis, or luteinization, and premature reproductive senescence. Uterine and vaginal weights and histopathology fluctuates with stage of the estrous cycle and pregnancy and therefore, high variability exists between cycling animals. As a consequence, only large treatment-related effects in weights for these tissues will be significant in the postnatal day 14 females in the OECD 421/422 studies. These tissues are estrogen-responsive and therefore, effects can range from atrophic tissue in response to disrupted steroidogenesis, to increased uterine weight and size due to exposure to an estrogenic compound. Finally, in both males and females, pituitary weight changes and associated histopathological changes signify adverse effects associated with an endocrine disruptive mechanism. Gonadotrophin-specific histopathological evaluation of the pituitary can help distinguish affected cell types and confirm changes associated with impaired reproductive function [77].

Overall, a great deal of information can be obtained from the OECD 421/422 reproductive/developmental screening studies including general toxicity assessments of nonpregnant, pregnant, and lactating adults, indications of reproductive or developmental toxicity, endocrine disruptor activity, and TK data through several life stages. Nevertheless, due to the limited sample size and duration of the study, biological significance of identified changes must be considered in the context of other information when available and in certain cases, interpreted as "equivocal" until additional testing is conducted. As stated in the guidelines "... this method will not provide evidence for definite claims of no reproduction/developmental effects. Moreover, in the absence of data from other reproduction/developmental toxicity tests, positive results are

useful for initial hazard assessment and contribute to decisions with respect to the necessity and timing of additional testing." In short, the OECD 421/422 screening studies can be powerful tools when used appropriately and when the utility of the assay is understood from the outset.

3.5 Two-Generation Reproductive Toxicity Study (OECD 416)

The two-generation reproductive toxicity study is conducted according to the OECD 416 test guideline and is designed to evaluate the potential effects of a test compound on male and female reproductive function, as well as survival, growth, and development of offspring.

In this study design, rats (recommended species) are exposed to the test article for approximately 10 weeks prior to breeding, and continuing through breeding (2 weeks), gestation (3 weeks), and until necropsy for each of two generations (Fig. 4). Each dose group (usually three) and control group should contain sufficient numbers of females (approximately 25) to achieve the guideline recommended number of pregnant females/group (i.e., 20) at each breeding phase. The Crl:CD(SD) strain of rat is used due to its high pregnancy rate, suitability for toxicity testing, and availability of historical control data. Oral (diet, drinking water, or gavage) administration is the recommended route of exposure specified in the relevant test guidelines. Other considerations can be made if the potential route of human exposure is via dermal or inhalation exposure. Per guideline, the parental (P) animals should be approximately 5–9 weeks of age at the initiation of treatment.

Dose level selection is based primarily on the results of a previously conducted reproduction and developmental toxicity screening study (OECD 421) and/or developmental toxicity probe study conducted preferably with the same strain and route of exposure. In addition, TK data on the parent compound or metabolites may be used in the dose level selection using the KMD dose level selection process outlined previously. From these previous toxicity data, the

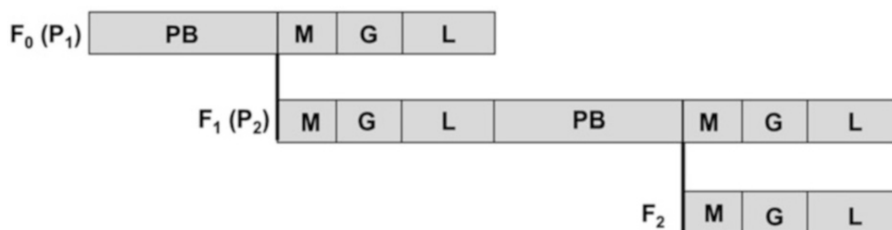

Fig. 4 Experimental design of the rat two-generation reproductive toxicity study. A control group and three treated groups are administered test material continuously for 10 weeks prior to breeding, throughout mating (2 weeks) and gestation (3 weeks), and up to the end of lactation (3 weeks) for each of two generations. In total, test material exposure is continuous for a total of approximately 42 weeks from the beginning of exposure in the first parental generation (P1) to the end of exposure in the second filial generation (F2). *PB* prebreeding, *M* mating, *G* gestation, *L* lactation

investigator selects a high dose (not to exceed the limit dose of 1000 mg/kg/day) with the aim of producing some parental toxicity or reproductive toxicity but not parental mortality exceeding 10 % or severe suffering of the animals. The remaining dose levels are usually spaced at three- to fourfold intervals and are selected to provide dose–response data for any observed treatment-related effects in the high dose group. The low dose is expected to be a NOAEL.

Study parameters and schedule for adults and their litters are summarized in Tables 4 and 5. All animals are monitored for growth and health throughout the study. The estrous cycle of the females is evaluated for length and normality for 3 weeks prior to mating, during cohabitation, and on the day of the scheduled necropsy. Females are observed for signs of parturition and any signs of difficulty or unusual duration are recorded. The day of parturition is recorded and the pups are examined. The following endpoints are recorded for each litter: the date of parturition; the number of live and dead pups; and the sex and body weight of each pup (before and after culling). Any abnormalities in the pups are recorded. To minimize variation in pup growth due to differences in litter size, litters may be standardized to eight pups per litter.

One male and one female per litter are randomly selected as P2 animals to produce the second generation. All F1 weanlings selected as the future P2 generation are observed daily for vaginal opening or for preputial separation. The age and body weight of the animals on the day of attainment are recorded. If a treatment-related effect on F1 sex ratio or age at pubertal onset is detected, anogenital distance (absolute and relative to the cube root of body weight) is measured in the F2 pups [78].

A complete necropsy with collection of organ weights and histopathological evaluation of reproductive and target organs is conducted on all adult animals/sex/generation. The histopathological examination of testis includes a qualitative assessment of stages of spermatogenesis. Examination of the ovaries includes enumeration of primordial follicles in the post-lactation P2 generation females. Sperm motility, spermatid/sperm epididymal and testicular counts and morphology are also evaluated in the adult males of each generation. A subset of weanlings is randomly selected for a gross pathological examination with or without the collection of organ weights.

3.5.1 Data Interpretation

The investigator interprets the results of the reproduction study for both parental and reproductive effects. Interpretation of parental effects may include effects on endpoints of clinical observations, growth (body weight/body weight gains), feed consumption, organ weights, and gross and histopathology of reproductive and target organs. Interpretation of reproductive effects may include

Table 4
Summary of two-generation study parameters and schedule (adults)

Study events and parameters	Number of animals	Timing (both generations unless indicated otherwise)
Cage-side examinations	All	At least twice daily
Clinical observations—males	All	Weekly throughout the study
Clinical observations—females	All	Weekly during pre-breeding period; GD 0, 7, 14, and 21; LD 1, 4, 7, 14, and 21
Body weights—females	All	Pre exposure, twice during the first week and weekly thereafter during pre-breeding period; GD 0, 7, 14, and 21; LD 1, 4, 7, 14, and 21
Body weights—males	All	Pre exposure, twice during the first week and weekly thereafter
Feed consumption—females	All	Twice during the first week and weekly thereafter during pre-breeding period; GD 0, 7, 14, and 21; LD 1, 4, 7, 11, 14, 17, 19, and 21
Feed consumption—males	All	Twice during the first week and weekly thereafter during pre-breeding period; weekly following breeding phase
Estrous cycle evaluation—females	All	Three weeks prior to mating; during mating; day of necropsy
Reproductive performance	All	N/A
Gross necropsy—adult females	All	After LD 21 or at least 24 days after evidence of mating or end of mating period for females not delivering a litter
Gross necropsy—adult males	All	After litters have been born
Organ weights (target and reproductive)	All	At necropsy
Sperm motility—males	All	At necropsy
Sperm count (testicular and epididymal)	All control and high-dose males	Post-necropsy
Sperm morphology	All control and high-dose males	Post-necropsy
Histopathology (including oocyte quantification) P2 only	All control and high-dose animals	Post-necropsy

GD gestation day, *LD* lactation day, *N/A* not applicable

effects on male and female reproductive function, as well as survival, growth, and development of the offspring.

The relationship to treatment of clinical observations or other incidence data is assessed by examining the overall incidence along

Table 5
Summary of two-generation study parameters and schedule (litters)

Study events and parameters	Number of animals	Timing (both generations unless indicated otherwise)
Dam/litter clinical observations	All	PND 0, 1, 4, 7, 14, and 21
Dam/litter cage-side observations	All	PND 0–21
No. of live and dead pups	All	PND 0, 1, 4, 7, 14, and 21
Pup sex and body weight	All	PND 1, 4 (BC and AC), 7, 14, and 21
Anogenital distance[a]	All	PND 1 (F2 pups only)
Culling	All	PND 4
Weaning	All	PND 21
Gross necropsy—weanlings	3/sex/litter	PND 22
Organ weights—weanlings	1/sex/litter	PND 22
Vaginal opening (VO)	1 female/litter	PND 28 until achieved (F1 only) or PND 43
Preputial separation (PPS)	1 female/litter	PND 35 until achieved (F1 only) or PND 53
Body weight at pubertal onset	All/sex/litter	Day of VO or PPS acquisition (F1 only)

PND postnatal day, *BC* before culling, *AC* after culling
[a]Anogenital distance is measured in F2 pups in the event of a treatment-related effect on F1 sex ratio or puberty onset

with other factors. Data interpretation often requires a holistic assessment of more than one endpoint due to the interrelatedness of endpoints examined in a reproduction study. For example, there may be an increased incidence of signs of dystocia (difficult birth) in the dams on a reproduction study that may be associated with an increase in gestation length. A treatment-related increase in the incidence of clinical signs consistent with pup mortality (dead, autolyzed, pale skin, placental tissue attached, cold to touch, and decreased activity and others) may also be noted and could be evident in conjunction with decreases in neonatal survival, pup body weights, and litter size. However, an increase in gestation length could also be associated with higher pup body weights. Both litter size and survival indices are important endpoints that may be impacted due to several factors and should be interpreted together with other reproductive endpoints.

A reduction in male or female fertility is another case where the data should be interpreted together with sex-dependent changes in either sperm parameters (motility, counts, or morphology), or ovarian follicle counts, and estrous cyclicity data along with histopathological evaluation.

Typically when assessing body weight or feed consumption data, examination of both the mean and individual animal data is needed

in order to determine the relationship to treatment of any higher or lower values. The magnitude and dose-relatedness of the values are also assessed. Whether or not the change is statistically identified or repeated in the second generation of the reproduction study is also taken into account. For example, if an endpoint was noted in the first generation of pups but not in the second generation, the investigator may conclude that the change in the first generation was spurious and of no toxicological significance. Recent laboratory historical control data of the same species/strain and route of exposure may be used to assess the normal variability of the endpoint as well as an indicator of changes that occur over time.

The investigator should understand the interaction of related endpoints. A simple example of this is a treatment-related decrease in body weight that is often associated with a concomitant decrease in feed consumption. The relationship of body weight to organ weights should also be considered. A decrease in an absolute organ weight may or may not be due to decreases in body weight; however, a significant decrease in body weight may impact the organ weight. A decrease or increase in weights of the reproductive organs may be associated with gross and histopathological changes and may be indicative of an effect on function or reproductive status. In general, changes in male or female reproductive organ weights should be interpreted together with reproductive performance and sex-dependent changes in either sperm parameters (motility, counts, or morphology), or ovarian follicle counts and estrous cyclicity data along with results from a histopathological evaluation.

Occasionally, a WoE approach may be needed to assess the relationship to treatment of a reproductive endpoint. For example, if there was an apparent treatment-related, statistically identified delay in pubertal onset for males in the high-dose group without a corresponding decrement in body weight at the time of attainment, a WoE approach may be considered when interpreting the data. While an apparent relationship to treatment for the delay in puberty onset cannot be ruled out, the investigator should determine whether or not this finding occurred in isolation. Since many factors contribute to puberty onset in male rats [79]; a WoE approach across androgen-sensitive endpoints may help determine whether or not there was a consistent pattern of altered androgenicity in the male treated rats. Using this approach, the investigator could consider anogenital distance data, male reproductive organ weight changes, gross or histopathological changes including qualitative testicular staging, and changes in sperm parameters (spermatid/sperm counts, sperm motility, and sperm morphology). A conclusion that there was no other indication of a change in androgen status could be drawn if the following occurred:

- There were no statistically identified or treatment-related effects on male anogenital distance in the F2 males. Anogenital distance is considered one of the most sensitive end points for altered androgen status [80].
- There was no evidence of hypospadias, undescended testis, or exposure-related testicular, epididymal, prostate, or seminal vesicle organ weights or histopathological changes. In addition, there was no effect on qualitative stages of spermatogenesis.
- There were no significant changes in spermatid/sperm counts, sperm motility, and sperm morphology.
- There were no treatment-related effects on male or female reproductive indices, including mating, fertility, time to mating, or gestation length.

Alternatively, the investigator may use a different approach when there is a delay in vaginal opening and/or preputial separation with a corresponding decrease in the body weight at age of attainment. The investigator first determines whether or not the values are within or outside the laboratory's historical control range. Secondly, the investigator determines if the maternal body weights are decreased due to test material exposure in the same treatment group. If decreased, the investigator may conclude that while treatment-related, these slight delays in vaginal opening or preputial separation were due to decrements in pup body weight secondary to lower maternal body weight and, therefore, of no toxicological significance. This conclusion is supported by feed restriction studies where pup body weight decrements greater than ~15 % at puberty attainment or weaning caused secondary delays in both preputial separation and vaginal patency [47, 81, 82].

Interpretation of changes in weanling organ weights may also take changes in body weight into consideration. For example, decreases in organ weights may be interpreted to be secondary to the lower body weights of these animals. The investigator should determine whether or not the pattern of organ weight alterations in the dose group are consistent with that seen in a one-generation feed restriction study in which organ weights in weanlings correlate closely with body weight [47]. The degree of change in the organ weights should also be considered. If the organ weight changes are minor and are associated with decreased body weight, the investigator could conclude that the organ weight changes were not considered to be toxicologically significant.

3.5.2 Future Directions

In 2011, a new OECD test guideline (OECD 443) for the EOGRTS was finalized that could potentially replace or be an alternative to the Two-Generation Reproductive Toxicity Study (OECD 416). This test guideline favorably aligns with the goal of the 3Rs (replace, refine, or reduce animal testing) and offers a tiered testing approach. In addition to the standard reproductive

endpoints offered in the two-generation reproductive toxicity study, this guideline also includes cohorts for both developmental neurotoxicity and developmental immunotoxicity. Furthermore, under this guideline, the second generation will be triggered based on the results of the first generation rather than required. On the other hand, drawbacks for this new testing approach include higher costs and laboratory management constraints. The following section describes the EOGRTS in detail.

3.6 Extended One-Generation Reproductive Toxicity Study (EOGRTS; OECD 443)

3.6.1 Regulatory Use of EOGRTS

The EOGRTS is a large, complex, highly integrated study design; therefore, careful consideration is warranted before this study type is selected. For agrochemicals, the EOGRTS may be selected during product reregistration to fill data gaps and/or update previously conducted registration studies. For example, if a pesticide active ingredient has an older two-generation reproductive toxicity study (i.e., conducted before the 1998 test guideline revision) and also requires additional registration data, such as a developmental neurotoxicity study (OECD 426) and/or endocrine assessments (e.g., a comparative thyroid study), then an EOGRTS may allow for the conduct of a single study instead of multiple studies and save animals, resources, and costs. For new pesticides, the studies required for pesticide registration will generate similar data across several studies to allow an evaluation of the same types of toxicity. In this case, an EOGRTS is not required.

For industrial chemicals, the scenario for using the EOGRTS design is different than agrochemicals. In 2015, amendments to annexes VII, IX and X of REACH [83], industrial chemicals produced at greater than ≥100 metric tons/year require a reproductive toxicity study (EOGRTS preferred) as part of their data set in accordance with REACH data requirements for reproductive toxicity. With lower volume chemicals, an EOGRTS may be triggered based on findings in other studies.

3.6.2 EOGRTS Range-Finding Study

Whether the study is conducted for an agricultural or industrial chemical, a range-finding study may be beneficial before conducting an EOGRTS. The decision to conduct a range-finding study will depend on several factors, including: (1) Whether there are available data that will allow the investigator to select dose levels in pregnant/lactating animals with some degree of confidence. If selected doses are too high, consequences could include nonspecific effects in the F1 offspring that may impact numerous toxicity endpoints or excessive toxicity resulting in insufficient offspring to fill the various cohorts and assess all endpoints needed in the EOGRTS design. Alternatively, the regulatory acceptability of a large, expensive EOGRTS study may be at risk if animals are not sufficiently challenged with the test compound. (2) Whether the investigator verifies lactational transfer. To gain full value from an EOGRTS

study, it is important to demonstrate that the pups were exposed to the test material during critical windows of development; otherwise, direct dosing of the pups should be considered. (3) Whether the investigator chooses to characterize toxicokinetics across life stages to set a KMD, thereby avoiding dose-dependent saturation of toxicokinetic processes and the subsequent development of irrelevant toxicity data. To characterize toxicokinetics, the test guideline recommends evaluating maternal and fetal/pup blood during late gestation (e.g., GD 20), during mid-lactation (e.g., LD 10) when pups are relying solely on maternal milk for nutrition, and shortly after weaning (e.g., postnatal day (PND) 28). Maternal milk also may be sampled on LD 10.

The range-finding study design may vary depending on data needs, and often a modified OECD 421/422 will serve this purpose. In one example, groups of 12 male and 12 female Crl:CD (SD) rats are fed diets supplying 0 (control), and three dose levels of a test substance. Males are exposed for at least 2 weeks prior to breeding and continuing throughout breeding and post-breeding for approximately 8 weeks of exposure. The females are exposed for 2 weeks prior to breeding, continuing through breeding (up to 2 weeks), gestation (3 weeks), and lactation (3 weeks). Effects on clinical observations, body weight/gain, gonadal function, mating behavior, conception, development of the conceptus, and parturition are evaluated. In the offspring, litter size, pup survival, sex, body weight, and the presence of gross external morphological alterations are assessed. In addition, as recommended in the OECD 443 test guideline, blood levels of the test substance can be assessed in P1 females on GD 20 and LD 10, and F1 offspring on GD 20, PND 10, and PND 28. Alternatively, lactational transfer could be evaluated in LD 4 dams and PND 4 cull pups to avoid using additional offspring. P1 males could be evaluated at termination. Test substance blood levels can be single point determinations or three samples/day for an AUC determination [19]. A limited gross necropsy and histopathology of reproductive and/or target organs can be conducted for P1 adults and selected F1 PND 21 offspring if this will aid in dose selection.

3.6.3 EOGRTS

The design for the EOGRTS is described in OECD test guideline 443, although there are some differences in study design depending on the regulatory program through which data collection is required. For pesticides, the study design described in the OECD 443 test guideline is applicable and additional guidance can be found in OECD Guidance Document No. 117 and 151. For industrial chemicals undergoing REACH registration, there are notable differences to the OECD 443 test guideline; these differences are outlined in the ECHA REACH information requirements as described in the “Guidance on Information Requirements and

Chemical Safety Assessment—Chapter R.7a: Endpoint Specific Guidance" [84]. These differences are described in greater detail below.

Exposures: For route of exposure, EOGRTS is most amenable to a dietary or drinking water route of exposure. Furthermore, these routes are preferred over oral gavage as more relevant to human to human exposures in many cases and for animal welfare reasons. While gavage also is an option, it is important to consider the number of F1 animals that may require gavage exposures during the peak of the study. If there are 20 litters/dose, four dose levels and litters are culled to eight pups, there are potentially 640 F1 animals that will require gavage administration of test material each day. The inhalation route also poses increased challenges for the EOGRTS design related to combining 6-h daily exposures with the necessary data collection while maintaining a daily 12-h light–dark cycle. During study planning, approximately 7 h/day are set aside to allow for the inhalation exposures, including loading and unloading exposure chambers, time to attain target chamber concentrations, and time to clear the chambers of test material after the exposures. This leaves 5 h/day for additional data collection, which may be problematic during some phases of the EOGRTS. For example, at time points around weaning (PND 14–28), study activities may include P1 clinical observations/detailed clinical observations, body weights, food consumption, litter data, F1 pup body weights, nipple retention, pup cohort assignments, pup weaning, necropsy of P1 dams and unselected F1 pups, auditory startle response in select F1 offspring, puberty onset evaluation in female offspring, and detailed clinical observations in select pups. Thus a well-organized and coordinated effort is needed to complete data collection and allow sufficient time in the exposure chambers without any impact on the 12-h light–dark cycle. Altered light–dark cycles or shifts in photoperiod have been associated with disrupted circadian profiles of plasma hormones, altered estrous cycles, altered diurnal locomotor activity, decreased maternal, fetal, and pup body weights, etc. [85–89]. These issues also are complex via the dermal route, where dermal wrapping can be difficult during breeding, gestation, and lactation, and would require an enormous amount of time to expose the F1 offspring. While the EOGRTS may be conducted by these alternate exposure routes, the study may need to be conducted in replicates and the resource requirements and logistics become more challenging. The remainder of this section describes EOGRTS using the dietary route of exposure.

For dose selection, the EOGRTS typically includes one control group and three treatment groups that are administered test material. The high dose can be selected using the traditional MTD approach or using a KMD approach (see above). If the EOGRTS is conducted for REACH, a KMD approach cannot be used; an

MTD approach must be used as hazard characterization is required for classification and labelling, regardless of the relationship to human exposure levels.

Once the high dose has been selected, two lower dose levels are selected which will provide information on dose–response relationships, derivation of a benchmark dose or identification of a NOAEL. During the EOGRTS, dietary concentrations may be adjusted during periods when food consumption/kg body weight is significantly elevated in order to maintain a more consistent "mg/kg/day" dose. This practice was justified in the original manuscript by Cooper et al. [90], which states: "When using fixed parts per million (ppm) dosages for dietary administration, the investigator should consider using ADME to support reducing the concentration of test substance during the lactation periods of the parental females and during the early-life stages of the F1 generation . . . the investigator should use ADME as well as best estimates of food consumption to avoid overdosing the animals during these critical life stages." Thus, laboratories may adjust dietary concentrations during the second and third week of lactation to account for large increases in food consumption. During the post-weaning period, pups also consume greater quantities of food/kg body weight for several weeks; thus, the longer the F1 animals are maintained on test diet beyond PND 70, the closer the test material intake values approached the targeted dose levels [29]. At earlier time points, F1 offspring will receive doses greater than the targeted dose levels.

Study Design: The EOGRTS study design is shown in Fig. 5. Group sizes in the EOGRTS are designed to achieve at least 20 pregnant females per dose group; typically, 25 male and 25 female CD rats/dose are used. P1 rats are fed diets supplying 0, low-, middle-, or high-dose of the test compound for approximately

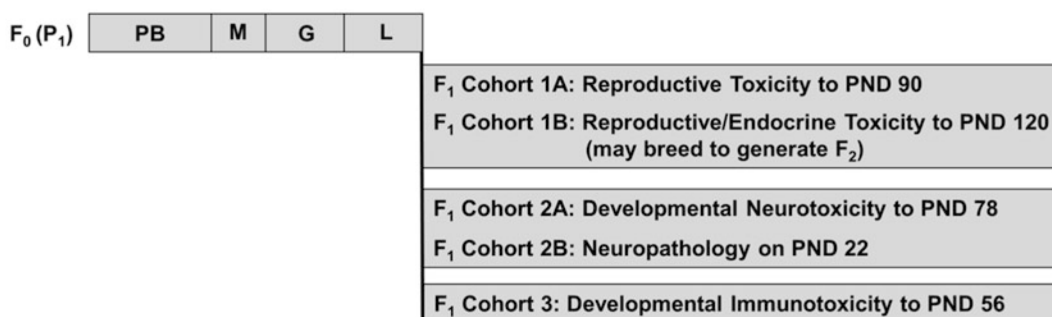

Fig. 5 Experimental design of the rat EOGRTS. A control group and three treated P1 groups (males and females) are administered test material continuously for 2 weeks prior to breeding, throughout mating (up to 2 weeks) and gestation (3 weeks), and up to the end of lactation (3 weeks). P1 males receive a minimum 10 week exposure. At weaning, F1 offspring are assigned to different cohorts for further DART testing. The F1 generation may be bred to produce an F2 generation, if needed. *PB* pre-breeding, *M* mating, *G* gestation, *L* lactation

2 weeks prior to breeding and continuing through breeding (up to 2 weeks), approximately six additional weeks (males receive a minimum 10-week exposure) or gestation (3 weeks) and lactation (3 weeks) for females in accordance with the OECD 443 test guideline. P1 females will be exposed until LD 22 (the end of the lactation period). Selected F1 offspring will be maintained on the test diet until PND 22 (Cohort 2B), PND 53–59 (Cohort 3), PND 78 (Cohort 2A), PND 90 (Cohort 1A), or PND ≥97 (Cohort 1B).

Note that if REACH compliance is required, P1 male and female rats must be exposed for 10 weeks during the prebreeding period to account for a full cycle of spermatogenesis/folliculogenesis prior to breeding. Subsequent exposures (i.e., mating, gestation, lactation, and the F1 offspring) are the same as outlined in the OECD 443 test guideline. REACH regulators have indicated that a 10-week prebreeding exposure is needed prior to the functional fertility assessment for "classification and labelling" purposes. However, analyses from previous studies indicate that the 10-week prebreeding period is not necessary to detect an effect on spermatogenesis or folliculogenesis, because more sensitive endpoints are included in the EOGRTS study design (e.g., reproductive organ histopathology, sperm motility, counts, and morphology) [59, 91–95].

One of the strengths of the EOGRTS is the inclusiveness of a breadth of toxicity endpoints to allow for an evaluation of reproductive, neurotoxicity, immunotoxicity, endocrine toxicity, and systemic toxicity following exposures during adulthood and during critical life stages. In-life parameters in P1 animals include clinical observations/detailed clinical observations, feed consumption, body weights, estrous cycle evaluation, litter and fertility data, thyroid hormone measurements, clinical chemistry/hematology parameters, and urinalysis. In addition, postmortem evaluations of P1 adults include gross pathology, numerous organ weights, extensive histopathology, and an evaluation of sperm parameters.

In-life parameters evaluated in all F1 offspring include clinical observations/detailed clinical observations, body weights, feed consumption, anogenital distance, nipple retention, and puberty onset. Depending on the specific EOGRTS design, F1 offspring may be divided into Cohorts 1A, 1B, 2A, 2B, and 3 at weaning (PND 21) as follows:

- Cohort 1A (≥20/sex/dose; 1 pup/sex/litter for 20 males and 20 females per dose) are used to evaluate reproductive/endocrine toxicity, including estrous cycle evaluation and postmortem evaluations focused on reproductive organs, sperm assessment, and ovarian follicle counts on PND 90. This group also is used to assess general systemic and thyroid toxicity, which includes clinical chemistry/hematology parameters, thyroid hormone assessment, and urinalysis. Postmortem evaluations

in Cohort 1A (PND 90) also include gross pathology, organ weights, and histopathology on a wide range of tissues (including thyroids) as well as an assessment of some immune parameters (e.g., lymph node weights, splenic lymphocyte enumeration).

- Cohort 1B animals (≥20/sex/dose; 1 pup/sex/litter for 20 males and 20 females per dose) are used to generate a second generation if needed (triggers for breeding a second generation are discussed below). If a second generation is not needed, this cohort is known as the endocrine group and is designated to clarify any equivocal responses seen in the Cohort 1A animals. Postmortem evaluations in Cohort 1B (≥PND 97) include gross pathology and organ weights with a primary focus on tissues affected in Cohort 1A, including thyroids.
- Cohort 2A (10/sex/dose; 1 pup/litter for ten males and ten females per dose) is used for developmental neurotoxicity (DNT) assessments, which includes functional observational battery (FOB), motor activity, and auditory startle response (ASR). Between PND 75 and 90, Cohort 2A F1 animals are perfused for central nervous system (CNS) and peripheral nerve neuropathology evaluation and brain morphometry. Typically, 12 animals/sex/dose are used in the event that there are any problems with brain morphometry samples (e.g., perfusion problem, missed brain landmark during sectioning) to ensure a sample size of 10/sex/dose for brain morphometry.
- Cohort 2B animals (10/sex/dose; 1 pup/litter for ten males and ten females per dose) undergo necropsy on PND 22, which includes brain weight collection in these weanlings and immersion fixation of tissues for examination of neuropathology.
- Cohort 3 animals (10/sex/dose) are used to assess potential developmental immunotoxicity using a T-cell dependent antibody response (TDAR) assay (e.g., sheep red blood cell antibody-forming cell assay) on PND 56 ± 3.

The endpoints evaluated in the EOGRTS are listed in Table 6.

Triggers for breeding a second generation: Data that trigger the production of a second generation are outlined in OECD Guidance Document No. 117 [96]. Basically, triggers include an adverse effect on: P1 fertility or fecundity, litter parameters, F1 survival, pup malformations, live birth index, F1 pup body weights, F1 developmental landmarks, or F1 estrous cyclicity, particularly when these effects cannot be attributed to severe maternal or systemic toxicity. In addition, if data in the concurrent control group are atypical (e.g., outside historical control ranges for the laboratory), breeding of a second generation may be useful. In considering whether to produce the F2 generation, it is important

Table 6
Primary EORGTS endpoints examined across different life stages

	P1 males	P1 females	F1 culled pups	Unselected F1 weanlings (non-perfused)	F1 Cohort 1A	F1 Cohort 1B	F1 Cohort 2A	F1 Cohort 2B	F1 Cohort 3
Sampling plan	>20/dose[a]	>20/dose[a]	Cull to 10: ~5/sex per litter	1 pup/sex / litter	1 pup/sex/litter	1 pup/sex/litter	1 pup/litter	1 pup/litter	1 pup/litter
N				10/sex/dose	≥20/sex/dose[b]	≥20/sex/dose	10–12/sex/dose	10/sex/dose	10/sex/dose
Age at necropsy	21 weeks[c]	LD 22	PND 4	PND 22	PND 90 ± 1	≥PND 97 ± 1	PND 78	PND 22	PND 56 ± 3
Blood collected									
TK									
Thyroid hormones[d]			X	X					
Clinical chemistry/ hematology	X	X			X				
Urinalysis	X				X				
Necropsy with gross exam				X				X	
Terminal body weights	X	X	X	X	X		X	X	X
Organ weights									
Adrenal	X	X		X	X				
Brain	X	X		X	X		X	X	X
Epididymides	X				X				
Heart									
Kidney	X	X		X	X				

Liver	X	X	X	X		
Ovaries		X		X		
Pituitary[b]	X	X	X	X		
Prostate	X			X		
Seminal vesicles[e]	X			X		
Spleen	X	X	X	X		X
Testes	X		X	X		
Thymus				X		X
Thyroid[f,g]	X	X	X	X		
Uterus		X		X		
Repro parameters						
Corpora Lutea						
Implantation Sites		X				
Resorptions		X				
Repro indices	X	X				
Litter parameters		X				
Devtl landmarks						
Anogenital distance[h]		X				
Nipple Retention[i]		X				
Puberty onset[j]				X	X	X
Sperm parameters[k]	X			X		
Histopath exams[l]						
Adrenal	X	X	X	X		

(continued)

Table 6
(continued)

	P1 males	P1 females	F1 culled pups	Unselected F1 weanlings (non-perfused)	F1 Cohort 1A	F1 Cohort 1B	F1 Cohort 2A	F1 Cohort 2B	F1 Cohort 3
Brain	X	X		X	X				X
Epididymides	X			X	X				
Heart					X				
Kidney	X	X		X*	X				
Liver	X	X		X*	X				
Lymph nodes					X				X*
Mammary gland	X	X			X				
Neuropathology[m]							X	X	
Ovaries[n]		X		X	X				
Pituitary	X	X		X	X				
Prostate	X				X				
Relevant gross lesions	X	X		X	X				
Seminal vesicles[e]	X				X				
Spleen	X	X		X	X				
Testes[o]	X			X	X				
Thymus					X				X*
Thyroid	X	X	X	X	X				
Uterus		X			X				

Quantitative brain assessment			
Gross brain measurements[p]	X	X	
brain morphometry	X		
Developmental immunotoxicity			
T-cell dependent antibody response			X
Natural Killer Cell Assay			

[a]10–12/sex examined histopathologically from control and high-dose groups; tissues with exposure-related changes are examined in the low- and middle-dose groups as well as reproductive organs from animals that failed to mate or are infertile

[b]10/sex/dose examined histopathologically except for testes and epididymides, which are examined in all control and high-dose males; tissues with exposure-related changes were examined in the low- and middle-dose groups as well as reproductive organs from animals that failed to mate or were infertile

[c]Necropsied after ~11 weeks of exposure

[d]T_3, T_4, and TSH

[e]With coagulating glands and fluid

[f]With parathyroid glands

[g]Post-fixation weights

[h]All pups

[i]All pups available after culling

[j]Data analyzed by litter in all F1 offspring

[k]Sperm/spermatid counts, sperm morphology, and sperm motility

[l]Most tissues fixed in neutral, phosphate-buffered 10 % formalin with the exception right testis, right epididymis, and ovaries, which are preserved in Bouin's fixative (also left testis and left epididymis unless used for assessing sperm parameters). Tissue sections are stained with hematoxylin and eosin unless otherwise specified

[m]Nine cross sections of the brain are examined including olfactory bulb, cerebrum (frontal, parietal, temporal, and occipital lobes), thalamus/hypothalamus, midbrain, pons, medulla oblongata, and cerebellum, as well as sections from the trigeminal ganglion and nerve, pituitary gland, eyes with optic nerves, spinal cord (cervical and lumbar), olfactory epithelium, and skeletal muscles (gastrocnemius and anterior tibial). In addition, sections of the cerebellum and corpus callosum are stained with Luxol Fast Blue to assess myelin. Spinal nerve roots (cervical and lumbar), dorsal root ganglia (cervical and lumbar), and peripheral nerves (sciatic, tibial (proximal and distal (muscular)—at the knee and calf muscle branches), and sural) are embedded in plastic and examined histopathologically

[n]Included ovarian follicle counts in females from the control and high-dose groups (15/dose)

[o]Testes stained with periodic acid Schiff's—hematoxylin

[p]Linear measurements include cerebral length and width, and cerebellar length and width

*Optional

to consider the limited parameter assessments in second generation animals, which are typically euthanized at weaning (i.e., evaluations could include reproductive performance, litter size, offspring survival and development (including anogenital distance and nipple retention), weanling organ weights and histopathology). Based on retrospective analyses of 498 previous multigeneration reproductive toxicity studies [97], the second generation seldom contributes critical data for hazard characterization or risk assessment; therefore, production of a second generation should be an infrequent occurrence. If toxicity triggers are limited to the high dose level alone (with no apparent dose-related trend), margin of exposure (MoE) between the effective dose level and estimated human exposures should be considered; if MoE is sufficient, an F2 generation may not be needed. The use of "in study" triggers requires that laboratories identify critical deadlines for data collection and analysis, and designate resources to complete these evaluations in a timely manner, so that important trigger information can be identified sufficiently early to permit a second breeding.

According to the REACH guidance document [84], ECHA would prefer that the decision on whether to breed a second generation is made prior to initiating an EOGRTS based on the following factors: (1) significant consumer exposure; (2) genotoxicity in somatic cell assays in vivo; (3) a protracted period (>1 week) to achieve steady state for internal dosimetry; and/or (4) evidence for a relevant endocrine MoA. Data from previous toxicity and TK studies are useful; however, the investigator may extend the EOGRTS if additional data generation would be beneficial to better understand toxicity.

Interpretation of EOGRTS data: The EOGRTS evaluates a large number of endpoints, many of which are interrelated; therefore, it is critical to use a WoE assessment when evaluating data from an EOGRTS. Questions to consider during data evaluation include: (1) Is there a pattern of effects across cohorts? (2) Is there a pattern of effects across generations? (3) Are related endpoints also affected? (although differential endpoint sensitivity also is important to consider) (4) Is there a dose–response relationship for the observed effects? (i.e., note that the nature of the effects may change with dose and non-monotonic dose-responses may occur, although likely to be relatively rare) (5) Do the results make sense in light of laboratory historical control data? (6) Could systemic toxicity contribute to the effects seen? (7) Are endpoints altered in the range of linear TK? This becomes particularly important if there is a large MoE between effective doses and estimated human exposures.

Examples of interpreting EOGRTS data in a WoE manner have been described for full and abridged EOGRTS designs [29, 98, 99]. One example is the interpretation of kidney effects in the 2,4-

dichlorophenoxyacetic acid (2,4-D) EOGRTS [29]. For 2,4-D, kidney was identified as the primary target organ for toxicity as demonstrated by very slight-to-slight histopathological changes in high-dose P1 males with marginally greater effects seen in high-dose F1 males and females. Thus, the reproducibility of kidney findings across generations, cohorts, and sexes improves confidence that this effect is treatment related. Interestingly, slight intergenerational differences in kidney toxicity were noted with 2,4-D exposure. Effects in the offspring may have greater incidence/magnitude in a given dose group due to enhanced sensitivity during critical windows of development, protracted periods of higher exposure that occur due to higher levels of food consumption per kg body weight, or a combination of these factors. Therefore, a careful comparison of exposures across life stages may be needed to determine which of these explanations is likely. In the case of 2,4-D, the slightly greater incidence and degree of kidney lesions in the F1 adults relative to the P1 generation was attributed to higher 2,4-D doses in the F1 offspring associated with nonlinear TK; it was not due to enhanced sensitivity of young animals to 2,4-D. This determination was facilitated by the well characterized dosimetry for kidney toxicity with 2,4-D.

Another example that indicates the importance of WoE is the finding of decreased seminal vesicle and prostate weights in the intermediate-dose and high-dose P1 males in the 2,4-D EOGRTS [29]. When interpreting these data, several points were considered: (1) absolute and relative organ weights of the concurrent P1 control group were atypical (greater than the laboratory HCD range), whereas the values for the treated animals were within the HCD; (2) there was no associated histopathology in these organs in 2,4-D-treated animals; (3) decreased accessory sex tissue organ weights were not reproduced in the F1 animals in three different cohorts, despite exposure to higher doses of 2,4-D for longer periods, which included critical windows of development; and (4) there were no effects on accessory sex tissue histopathology or sperm parameters in the 2,4-D-treated F1 offspring. Thus, this effect was not attributed to 2,4-D treatment. Inherent variability in male accessory sex tissue weights may have contributed to the group differences [100].

Systemic toxicity can affect multiple endpoints evaluated in the EOGRTS. Maternal body weight gains during gestation can affect pup body weights and weanling organ weights. With greater effects on body weight, F1 pups may exhibit altered anogenital distance, delayed puberty onset, altered estrous cycles, and decreased epididymal sperm counts [47, 101]. Furthermore, effects may occur secondary to other toxicities; for example, kidney toxicity (e.g., hypovolemia and electrolyte imbalance) can alter endocrine endpoints like adrenal gland histopathology (e.g., hypertrophy of the

zona glomerulosa to produce aldosterone) [102, 103]. Systemic toxicity is well characterized in the EOGRTS design, which allows for a critical evaluation of all data to better understand primary vs. secondary treatment-related toxicities.

With so many endpoints evaluated in an EOGRTS study and a 5 % chance for Type II errors ($\alpha \leq 0.05$), there is a likelihood that some statistically identified values are not toxicologically meaningful. If equivocal data occur during conduct of the EOGRTS, the reproducibility across cohorts can be used to discredit or confirm treatment-related findings, further characterization of an effect, or to provide information on whether an effect is sustained into adulthood.

3.6.4 Advantages of the EOGRTS Study Design

The EOGRTS study provides an integrated evaluation of systemic toxicity, reproductive toxicity, developmental neurotoxicity, developmental immunotoxicity, and endocrine toxicity in animals exposed during critical windows of development. Thus, a single study allows for a comparison of multiple endpoints and the dose levels at which toxic effects occur to determine the most sensitive target organ/effect and whether effects may be interrelated. The F1 animals are exposed during critical windows of development (gestation, lactation through adulthood), which allows for a determination as to whether animals may be more sensitive to toxic effects during critical life stages and which systems may be affected. A comparison of the endpoints in the EOGRTS test guideline (OECD 443) and other test guidelines/study guidance documents are shown in Tables 7, 8, 9, 10, and 11.

Another advantage of the EOGRTS is that many endpoints are assessed in both the P1 and F1 cohorts (estrous cyclicity, thyroid endpoints, sperm parameters, etc.) or in multiple F1 cohorts (puberty onset, body/organ weights, histopathology). These replicate findings increase the accuracy and interpretive value of the EOGRTS. Furthermore, the EOGRTS involves a more thorough assessment of endocrine endpoints relative to other available study designs.

The EOGRTS examines up to four pups/sex/litter beyond weaning, depending on which cohorts are included in the study design. Thus, more F1 offspring are evaluated than in a two-generation reproductive toxicity study, which only examines one pup/sex/litter into adulthood. This improves the precision of data collected across cohorts and improves the likelihood of detecting low incidence effects if they occur.

If the second generation is not produced, the EOGRTS is a shorter duration study (~21–29 weeks depending on a 2-week or 10-week pre-breeding period compared with ~42 weeks for a two-generation study) and the EOGRTS uses considerably fewer animals (i.e., ~1200 fewer animals) than the two-generation

Table 7
Repeated dose 90-day study endpoints[a]

Endpoint	Study type: Repeated dose 90-day Oral Tox (OECD 408)	Study type: EOGRTS (OECD 443)[b]
Dosing duration	~13 weeks (90 days)	P1 Adults: ♂ ≥13–15 weeks; ♀ - 16–18 weeks; F1: in utero, lactation, +10 week (1A to PND 90) or +≥11 week (1B to ~PND 98 or longer)
Age during dosing	Dosing to begin ASAP after weaning; before 9 weeks old	P1: 5–9 weeks of age (similar to timing in two gens with the longer pre-breeding exposure period) F1: Birth to PND 90 (1A) or ≥98 (1B) (if no F2 generation)
No. of animals	10/sex/dose (satellite: add 5/sex/group in control and high dose for reversibility)	P1: ≥20/sex/dose F1: ≥20/sex/dose (1A and 1B); 10/sex/dose (2 and 3)
Body weight (BWt)	At least weekly and terminal BWt	P1 weekly BWts with designated times in ♀ during gestation/lactation); F1 BWts on designated days prior to weaning and weekly thereafter; terminal BWts (all)
Feed consumption	At least weekly	P1 and F1: At least weekly
Clinical observations	1×/day	P1 and F1: 1×/day
Detailed clinical observations (DCOs)	1×/week	All P1 adults and post-weaning F1 offspring (1×/weekly)
Ophthalmological exam	Prior to exposure and at termination	P1 and F1 (1A): Histopathology on eye plus optic nerve
Motor and sensory function[c]	10/sex/dose	10/sex/dose (Cohort 2A, if included) (also, landing foot splay and auditory startle)
Hematology[d]	10/sex/dose (at termination)	P1: 10/sex/dose (at termination) Partial[e] (F1 only) or full scale: F1 (1A): 10/sex/dose
Clinical chemistry[f]	10/sex/dose (at termination)	Partial[g] (P1 only) or full scale: P1: 10/sex/dose (plus T4 and TSH) (at termination) F1 (1A): 10/sex/dose (plus T4 and TSH)
Urinalysis[h]	10/sex/dose (last week of study)	P1: 10/sex/dose (last week of study or end of pre-breeding) F1 (1A): 10/sex/dose (1A)
Gross necropsy	10/sex/dose	P1: ≥20/sex/dose F1 (1A and 1B): ≥20/sex/dose F1 (2A, 2B, 3): 10/sex/dose

(continued)

Table 7 (continued)

Endpoint	Study type: Repeated dose 90-day Oral Tox (OECD 408)	Study type: EOGRTS (OECD 443)[b]
Organ Wts	10/sex/dose (liver, kidneys, adrenals, testes, epididymides, uterus, ovaries, thymus, spleen, brain, and heart)	P1 and F1 (1A): ≥20/sex/dose (liver, kidneys, adrenals, testes, epididymides, uterus, ovaries, thymus, spleen, brain, heart, prostate, seminal vesicles, pituitary, thyroid and target tissues); terminal vaginal smear in all P1 and F1 females for estrous stage F1 (1A; 10/sex/group): local and distal lymph nodes Some organ Wts are reevaluated in F1 (1B)
Histopathology[i]	10/sex/dose (control and high dose; lower doses as needed) (gross lesions, brain—three sections, pituitary, thyroid, parathyroid, thymus, salivary glands, esophagus, stomach, small and large intestines w/Peyer's patches, liver, pancreas, kidneys, adrenals, spleen, heart, trachea, lungs, aorta, gonads, uterus, accessory sex organs, female mammary gland, prostate, urinary bladder, peripheral nerve (sciatic or tibial), bone marrow, skin, eyes (if changed during ophthalmology exam) and lymph nodes (local and distal)	P1 and F1 (1A): ≥20/sex/dose (control and high dose; lower doses as needed) (target organs, gross lesions brain—nine sections, pituitary, thyroid, parathyroid, thymus, gi tract (esophagus, stomach, small and large intestines w/Peyer's patches), liver, kidneys, adrenals, spleen, heart, trachea, lungs, gonads (includes ovarian follicle and corpora lutea counts in F1 Cohort 1A), uterus, vagina, accessory sex organs, vas deferens, female mammary gland, male mammary gland, prostate, urinary bladder, peripheral nerve (sciatic or tibial), muscle, spinal cord, eye plus optic nerve, bone marrow, and lymph nodes (local and distal—F1 (1A) 10/sex/dose); Histopathology also may be conducted on F1(1B) offspring if warranted
Satellite Group	5/sex/dose (evaluate recovery of any treatment-related findings)	Satellite group is not included in TG

[a]Underlined endpoints are identified in one test guideline, but not the other test guideline

[b]As outlined, the EOGRTS focuses on the P1 and F1 (1A and 1B) animals and does not include endpoints evaluated in other cohorts with the exception of neurobehavioral endpoints evaluated in Cohort 2A

[c]Functional testing: hand-held and open-field observations, sensory evaluation [pupil response, response to touch and sharp noise, muscle tone (with extensor thrust response), nociception (tail pinch)], forelimb and hindlimb grip performance, body temperature

[d]Hematocrit, hemoglobin conc., erythrocytes count, total and differential leukocyte count, platelet count and blood clotting time/potential

[e]Partial hematology does not include hematocrit, hemoglobin conc., platelet count and blood clotting time/potential

[f]Sodium, potassium, glucose, total cholesterol, urea, blood urea nitrogen, creatinine, total protein, albumin, and >two enzymes (e.g., alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase, gamma glutamyl transpeptidase, and sorbitol dehydrogenase; bile acids optional)

[g]Partial clinical chemistry does not include sodium, potassium, blood urea nitrogen (although urea is included), and only requires two enzymes (not more than two); however, clinical chemistry in EOGRTS requires serum T4 and TSH levels

[h]Appearance, volume, osmolality or specific gravity, pH, protein, glucose and blood/blood cells (cell debris also included in EOGRTS)

[i]In both studies, histopathology is so extensive that laboratories may default to "full histopathology" for both study types

Table 8
Reproductive toxicity endpoints

Endpoint	Study type: 2-Generation reproduction toxicity study (OECD 416)	Study type: EOGRTS (OECD 443)
Dosing duration	16–18 weeks (P1); **26–28 weeks (F1/ P2); 6 weeks (F2)** (2 generations)[a]	16–18 weeks (P1); 17–20 week (PND 98+ in F1)[a]
Anogenital distance	**Triggered (F2 Offspring)**	Required (All F1 Offspring)
Puberty onset	**1 pup/sex/litter**	3-4 pups/sex/litter
Estrous cyclicity	20/dose (2 gens) **(3 weeks)**	20/dose (2 gens) (2 weeks + after VO)
Precoital indices	20/sex/dose **(2 gens)**	20/sex/dose (1 gen unless triggered)
Mating and fertility indices	20/sex/dose **(2 gens)**	20/sex/dose (1 gen unless triggered)
Implantation data	20/sex/dose **(2 gens)**	20/sex/dose (1 gen unless triggered)
Postimplantation loss	20/sex/dose **(2 gens)**	20/sex/dose (1 gen unless triggered)
Gestation length	20/sex/dose **(2 gens)**	20/sex/dose (1 gen unless triggered)
Offspring survival/ sex ratio	20/sex/dose **(2 gens)**	20/sex/dose (1 gen unless triggered)
Litter size	20 litters **(2 gens)**	20 litters (1 gen unless triggered)
Epididymal sperm count	20/dose (control and high; 2 gens)	20/dose (control and high; 2 gens)
Testicular spermatid count	**20/dose (control and high; 2 gens)**	Not required
Sperm morphology	20/dose (control and high; 2 gens)	20/dose (control and high; 2 gens)
Sperm motility	20/dose (2 gens)	20/dose (2 gens)
Ovarian follicle counts	F1 Adult Females **(subset)**	F1 Adult Females (F1 Cohort 1A)
Nipple retention (males)	**Not Required**	All F1 Offspring
Reproductive organ weights	**20/sex/dose (2 gens)**	20/sex/dose (P); 40/sex/dose (F1 Cohorts 1A and 1B)
Reproductive organ histopath	20/sex/dose (2 gens)	20/sex/dose (P and F1 Cohort 1A; examine F1 Cohort 1B if needed)

Bold text indicates differences in the 2-generation reproduction toxicity study (OECD 416) and the 2,4-D EOGRTS study compared with the EOGRTS (OECD 443) test guideline

[a]Includes in utero and lactational exposure period (6 weeks) in offspring

Table 9
Developmental neurotoxicity endpoints

DNT endpoint	Study type: Developmental neurotoxicity guideline (OECDE 426)	Study type: EOGRTS (OECD 443)
Dosing duration for offspring examined for DNT endpoints	GD 6—LD 21	Continuous exposure to F1 across all life stages (including in utero and postnatal)
Detailed clinical observations (DCOs) in parental generation	≥10 dams/dose (2-gestation; 2-lactation)	All P1 adults (weekly)
F1 DCOs	1 pup/sex/litter (20/sex/dose)	Weekly in F1 offspring post-weaning
Auditory startle	1 pup/sex/litter (20/sex/dose)	10/sex/dose (PND 24 ± 1)
Puberty onset	1 pup/sex/litter (20/sex/dose)	3–4 pups/sex/litter
Neuropath/brain morphometry (PND 22)	1 pup/litter (10/sex/dose)	Neuropathology only (10/sex/dose)
Behavioral ontogeny (surface righting)	1 pup/sex/litter (20/sex/dose)	Not required
Motor activity (PND 13, 17, 21)	1 pup/sex/litter (20/sex/dose)	Not required
Motor activity (adult)	1 pup/sex/litter (20/sex/dose)	10/sex/dose (PND 63–75)
Motor and sensory function (adolescent)[a]	1 pup/sex/litter (20/sex/dose)	Not required
Motor and sensory function (adult)[a]	1 pup/sex/litter (20/sex/dose)	10/sex/dose (PND 63–75)
Learning and memory (adolescent)	1 pup/litter (10/sex/dose)[b]	Not required
Learning and memory (adult)	1 pup/litter (10/sex/dose)[b]	Not required
Neuropath/brain morphometry (adult)	1 pup/litter (10/sex/dose)	1 pup/litter (10/sex/dose)

[a]FOB: Landing foot splay, forelimb and hindlimb grip performance, extensor thrust response, tail pinch, body temperature
[b]Larger sample sizes could be considered, e.g., up to 1 pup/sex/litter

reproductive toxicity study. The savings in animals is greater when one considers the additional study designs that can be omitted when conducting an EOGRTS with all cohorts included (Table 12).

Lastly, in cases where TK data are used, the EOGRTS allows for the use of a KMD approach to dose selection. This ensures that marked TK saturation does not occur. TK saturation can affect

Table 10
Developmental immunotoxicity endpoints

Endpoint	Study type	
	Immunotoxicity (870.7800)	EOGRTS (OECD 443)
Dosing duration	≥28-days	Continuous exposure to F1 across all life stages (including in utero and postnatal)
T cell-dependent antibody response	≥8 animals/dose; one sex only	10/sex/dose (Cohort 3)
Positive control group (TDAR)	Required	Required or justify if omitted
Enumeration of splenic lymphocyte subpopulations	Optional (consider if TDAR+)	10/sex/dose (Cohort 1A)
Splenic natural killer cell assay	Optional (consider if TDAR−)	Not required
Spleen wts	≥8 animals/dose	10/sex/dose (weanlings); 20/sex/dose (2 gens)
Thymus wts	≥8 animals/dose	10/sex/dose (weanlings); 20/sex/dose (2 gens)
Lymph node wts	Not required	10/sex/dose (Cohort 1A)
Spleen histopathology	Not required	20/sex/dose (2 gens); 10/sex/dose (weanlings - optional)
Thymus histopathology	Not required	20/sex/dose (2 gens); 10/sex/dose (weanlings - optional)
Lymph node, bone marrow histopath	Not required	10/sex/dose (Cohort 1A)

absorption, internal dose, tissue distribution, and rate of clearance, thereby altering the toxicity of the test compound [20, 29]. If humans are not exposed to doses that result in TK saturation, these parameters would not be affected; in these cases, a KMD approach produces data that are more relevant for realistic hazard characterization and risk assessment purposes.

3.6.5 Disadvantages of the EOGRTS Study Design

Many of the disadvantages of the EOGRTS design have been discussed previously. Unless a second generation is produced, the EOGRTS evaluates reproductive behavior and performance in the one generation (P1) only. Furthermore, P1 animals are not exposed for a full cycle of spermatogenesis or folliculogenesis prior to breeding, although sensitive evaluations of these processes are included in the EOGRTS in animals that have been exposed to test compound for the full cycles.

From a logistical perspective, the EOGRTS is a complex study design, which may be difficult to interpret. If "in study" triggers are used, data collection and analysis must be timely to determine

Table 11
Comparative thyroid endpoints

Endpoint	Study type	
	Comparative thyroid guidance	**EOGRTS (OECD 443)**
Dosing duration	GD 6—LD 21	Continuous exposure to P1 and F1 sets across all life stages
Pregnant dam thyroid evaluation[a]	10 dams/dose	Not required
Fetal thyroid evaluation[a]	10 litters/dose	Not required
Cull pup thyroid evaluation[a]	20 litters/dose	Thyroid hormones optional[b]
Adult/lactating dam thyroid evaluation[a]	20 dams/dose	P ♂ and ♀ (weights and histopathology; 20/sex/dose)
Weanling thyroid evaluation[a]	20 litters/dose (1♂ and 1♀/litter)	Unselected F1 offspring (PND 21–22; 10/sex/dose)
Adult F1 thyroid evaluation[a]	Not required	Set 1a F1 offspring (~PND 90; hormones –10/sex/dose; weights and histopathology—20/sex/dose)

[a]Thyroid evaluation includes thyroid hormones (TSH, T4 and sometimes T3), thyroid weight and thyroid histopathology unless otherwise specified
[b]Blood can be pooled by litters

Table 12
Animal usage for EOGRTS vs. other guideline studies

Study type	Adults[a]	Fetuses/pups[b]	Totals	Grand total
EOGRTS	200	~1300	~1500	
				~1500
Two-generation toxicity	200	~2600	~2800	
Developmental neurotoxicity (DNT)	100	~1300	~1400	
Comparative thyroid study (CTS)	160	~2080	~2240	
				~6440

[a]Assumes 25 males and 25 females/dose group to achieve 20 pregnancies/dose group (EOGRTS, Two-gen), 25 time-mated females/dose group (DNT) and 40 time-mated females/dose group (CTS)
[b]Assumes 13 pups/litter

whether a breeding trigger is met and a second generation is needed. Lastly, the EOGRTS is resource intensive and expensive, costing 2–3× more than a two-generation study, depending on the number of F1 cohorts included.

3.6.6 Challenges When Conducting an EOGRTS

Most of the endpoints included in the EOGRTS are routinely conducted by toxicology testing laboratories; however, the size and complexity of the EOGRTS study can make study logistics and data collection challenging. To begin, the F1 endpoints in the EOGRTS are typically driven by the age of the offspring, so each animal is evaluated for specific endpoints at a particular age/ age range. Given that animals are allowed a 2-week breeding period, the age of the animals within a cohort can vary by up to 2 weeks; therefore, there are multiple collection days for many of the endpoints evaluated in the EOGRTS. This can become particularly complex when there are a myriad of endpoints evaluated in different litters on the same study day. As noted above, at the time of weaning (PND 21 ± 7 days), laboratories may be collecting the following data on the same day in different animals/litters: P1 clinical observations/detailed clinical observations, body weights, food consumption, litter data, F1 pup body weights, nipple retention, pup cohort assignments, pup weaning, necropsy of P1 dams and unselected F1 pups, auditory startle response in select F1 offspring, puberty onset evaluation in female offspring, and detailed clinical observations in select pups. The time around weaning of the F1 offspring is a particularly busy period, so laboratories must be very organized to ensure that no data collection or cohort assignments are missed.

It is sometimes difficult to fill all cohorts with the available F1 offspring. If litters have only limited number of pups of one sex, then pups from that litter must be carefully assigned to enable that litter to be represented in all cohorts. This requires that pup cohort assignment is prioritized. Thus, pups are first assigned to cohort 1A, then cohort 1B to ensure that sample sizes are achieved for primary study endpoints. If the litter is limited (e.g., only three males or only three females), then one of each sex is assigned to cohort 2A and 2B and the sex represented by most pups in the litter (e.g., remaining females or males) is assigned to cohort 3. This means that the sex balance in cohort 3 will have to be balanced with other litters at that dose level. If there are one or two pups of one sex in a litter, that litter cannot be represented in all cohorts if a full cohort study is conducted (i.e., Cohorts 1A, 1B, 2A, 2B, and 3). Pup assignment is further complicated by the birth of litters over a 2-week period, so the full picture of pup assignments may not be available until the first litter has reached LD/PND 14.

In addition, the assignment of pups to cohorts based on birthdate and litter of origin does not allow for the randomization of pups by body weight at weaning. This can result in artifactual differences in body weight (and sometimes organ weights) across cohorts at this life stage. This must be considered when interpreting these data; if there is a treatment-related effect on body weights, a consistent pattern is likely across cohorts.

The staggered delivery of litters over a 2-week period requires that some endpoints are collected over multiple days when a group of animals achieves the appropriate age/age range. The EOGRTS is an integrated study design across multiple disciplines, many of which are not familiar with staggered data collection and variable group sizes on different collection days. For these disciplines, it is important to review when and how many animals will reach the proper age for testing, necropsy, etc. It is important to consolidate data collection to the extent possible to limit interday variability (e.g., test across an age range when appropriate); however, it must be recognized that some data (e.g., organ weight variability) may be affected by allowing use of an age range.

As stated above, it is important that data are frequently reviewed for triggers that require production of the F2 generation. In some cases, this is not difficult, as reproductive parameters in the P1 generation are collected and can be easily evaluated before the F1 Cohort 1B animals reach PND 97, when a decision on necropsy or breeding must be made. However, other triggering endpoints like F1 estrous cyclicity, which is evaluated for 2 weeks beginning around PND 75, may be difficult to complete before a decision on breeding the second generation is needed (e.g., some Cohort 1B females will reach PND 97 before the 2-week estrous cycle evaluation has been completed in all Cohort 1B females; therefore, it is likely that the decision to necropsy or breed the Cohort 1B animals may be made after PND 97). In any event, researchers must make a concerted effort to stay current on data collection and review so that these decisions can be made in a timely manner.

4 Conclusion

This chapter is intended to serve as a DART testing knowledge resource for new as well as experienced laboratory personnel. Although DART testing guidelines were established many years ago, the guidelines and practices continue to evolve as new scientific information is obtained (e.g., anogenital distance as a new endocrine endpoint and the KMD approach to dose-level selection). Thus, this chapter represents a snapshot in time regarding DART testing methodology; however, personnel performing DART testing should be cognizant of best practices and strive to incorporate new scientific information (where appropriate) into DART study designs.

Acknowledgements

The authors would like to thank Dr. Lynea Murphy for critically reading the manuscript.

References

1. European Commission (2006) Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Off J Eur Union, L396/1 of 30.12.2006. Commission of the European Communities. Available at: http://eur-lex.europa.eu/legal-content/en/TXT/PDF/?uri=CELEX:02006R1907-20160203. Accessed 4 Apr 2016
2. European Commission (2004) Regulation (EC) No 1935/2004 of the European Parliament and of the Council of 27 October 2004 on materials and articles intended to come into contact with food and repealing Directives 80/590/EEC and 89/109/EEC. Available at: http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:02004R1935-20090807. Accessed 4 Apr 2016
3. Responsible Care (2015) Available from: http://www.icca-chem.org/en/Home/Responsible-care/. Cited 4 Apr 2016
4. Provisions on Environmental Administration of New Chemical Substances (2010) M.o.E.P.o. China, Editor
5. Federal Insecticide, Fungicide, and Rodenticide Act (1996) (7 U.S.C. 136 et seq.)
6. United Nations (2015) Globally harmonized system of classification and labelling of chemicals (GHS), 6th revised edn
7. Sparks TC (2013) Insecticide discovery: an evaluation and analysis. Pestic Biochem Physiol 107(1):8–17
8. Wu S et al (2013) Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chem Res Toxicol 26(12):1840–1861
9. Ellis-Hutchings RG et al (2012) In vitro screening methods for developmental toxicology. In: Hood RD (ed) Developmental and reproductive toxicology. A practical approach. Informa Healthcare, New York
10. Palmer JA et al (2013) Establishment and assessment of a new human embryonic stem cell-based biomarker assay for developmental toxicity screening. Birth Defects Res B Dev Reprod Toxicol 98(4):343–363
11. Boobis AR et al (2006) IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit Rev Toxicol 36(10):781–792
12. Sonich-Mullin C et al (2001) IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Regul Toxicol Pharmacol 34(2):146–152
13. Seed J et al (2005) Overview: using mode of action and life stage information to evaluate the human relevance of animal toxicity data. Crit Rev Toxicol 35(8–9):664–672
14. Meek ME et al (2003) A framework for human relevance analysis of information on carcinogenic modes of action. Crit Rev Toxicol 33(6):591–653
15. ECHA (2014) Guidance on information requirements and chemical safety assessment, Chapter R.7c: endpoint specific guidance, version 2.0. European Chemicals Agency, Helsinki
16. Russell WMS, Birch RL (1959) The principles of humane experimental technique. Methuen, London
17. Terry C et al (2014) Application of a novel integrated toxicity testing strategy incorporating “3R” principles of animal research to evaluate the safety of a new agrochemical sulfoxaflor. Crit Rev Toxicol 44(Suppl 2):1–14
18. Saghir SA et al (2012) Assessment of diurnal systemic dose of agrochemicals in regulatory toxicity testing—an integrated approach without additional animal use. Regul Toxicol Pharmacol 63(2):321–332
19. Saghir SA et al (2006) Strategies to assess systemic exposure of chemicals in subchronic/chronic diet and drinking water studies. Toxicol Appl Pharmacol 211(3):245–260

20. Saghir SA et al (2013) Life-stage-, sex-, and dose-dependent dietary toxicokinetics and relationship to toxicity of 2,4-dichlorophenoxyacetic acid (2,4-D) in rats: implications for toxicity test dose selection, design, and interpretation. Toxicol Sci 136 (2):294–307
21. Jakubowski M, Terkel J (1985) Incidence of pup killing and parental behavior in virgin female and male rats (Rattus norvegicus): differences between Wistar and Sprague–Dawley stocks. J Comp Psychol 99(1):93–97
22. Creton S et al (2012) Use of toxicokinetics to support chemical evaluation: Informing high dose selection and study interpretation. Regul Toxicol Pharmacol 62(2):241–247
23. Jochemsen R et al (1993) Assessment of drug exposure in rat dietary studies. Xenobiotica 23 (10):1145–1154
24. OECD (2011) Test guideline 416: Two-generation reproduction toxicity study. OECD, Paris
25. OECD (2011) Test guideline 443: extended one-generation reproduction toxicity study. OECD, Paris
26. OECD (2011) Guidance document 116 on the conduct and design of chronic toxicity and carcinogenicity studies, supporting test guidelines 451, 452 and 453. OECD, Paris
27. Barton HA et al (2006) The acquisition and application of absorption, distribution, metabolism, and excretion (ADME) data in agricultural chemical safety assessments. Crit Rev Toxicol 36(1):9–35
28. Bessems JG, Geraets L (2013) Proper knowledge on toxicokinetics improves human hazard testing and subsequent health risk characterisation. A case study approach. Regul Toxicol Pharmacol 67(3):325–334
29. Marty MS et al (2013) An F1-Extended One-Generation Reproductive Toxicity Study in Crl:CD(SD) rats with 2,4-dichlorophenoxyacetic acid. Toxicol Sci 136 (2):527–547
30. Gidenne T, Lebas F, Fortun-Lamothe L (2010) Feeding behavior of rabbits. In: de Blas C, Wiseman J (eds) Nutrition of the rabbit. The Centre for Agriculture and Bioscience International, pp 233–254
31. Kopf R, Lorenz D, Salewski E (1964) Naunyn Schmiedebergs [The effect of thalidomide on the fertility of rats in reproduction experiments over 2 generations]. Arch Exp Pathol Pharmakol 247:121–135
32. Wilson JG (1953) Influence of severe hemorrhagic anemia during pregnancy on development of the offspring in the rat. Proc Soc Exp Biol Med 84(1):66–69
33. Aliverti V et al (1979) The extent of fetal ossification as an index of delayed development in teratogenic studies on the rat. Teratology 20(2):237–242
34. Glockner R, Schwarz S, Jahne F (1991) Simple methods for the characterization of skeletal development of rat pups during the perinatal period as a parameter of maturity. Anat Anz 173(4):209–214
35. Staples R (1974) Detection of visceral alterations in mammalian fetuses. Teratology 9:37
36. Wilson J (1965) Embryological considerations in teratology. In: Warkany J, Wilson JG (eds) Teratology: principles and techniques. University of Chicago Press, Chicago, p 251
37. Kang YJ, Zolna L, Manson JM (1986) Strain differences in response of Sprague–Dawley and Long Evans Hooded rats to the teratogen nitrofen. Teratology 34(2):213–223
38. Tyl R, Marr M (2012) Developmental toxicity testing-methodology. In: Hood R (ed) Developmental and reproductive toxicology: a practical approach, 3rd edn. Informa Healthcare, London and New York, p 167
39. Stuckhardt JL, Poppe SM (1984) Fresh visceral examination of rat and rabbit fetuses used in teratogenicity testing. Teratog Carcinog Mutagen 4(2):181–188
40. Trueman D, Jackson SW, Trueman B (1999) An automated technique for double staining rat and rabbit fetal skeletal specimens to differentiate bone and cartilage. Biotech Histochem 74:98–104
41. Zablotny C (2002) Thermo Shandon's Pathcentre: a tissue processor or a teratology processor? Lab Reader Thermo Shandon, pp 1–3
42. Dawson A (1926) A note on staining of the skeleton of cleared specimens with Alizarin Red-S. Stain Tech 1:1–21
43. US EPA (1991) Guidelines for developmental toxicity risk assessment. In: Register F (ed), vol 56
44. Beyer BK et al (2011) ILSI/HESI maternal toxicity workshop summary: maternal toxicity and its impact on study design and data interpretation. Birth Defects Res B Dev Reprod Toxicol 92(1):36–51
45. Giavini E, Menegola E (2012) The problem of maternal toxicity in developmental toxicity studies. Regul Toxicol Pharmacol 62 (3):568–570
46. ECHA (2013) Guidance on the application of the CLP criteria

47. Carney EW et al (2004) The effects of feed restriction during in utero and postnatal development in rats. Toxicol Sci 82 (1):237–249
48. Fleeman TL et al (2005) The effects of feed restriction during organogenesis on embryo-fetal development in the rat. Birth Defects Res B Dev Reprod Toxicol 74(5):442–449
49. Garofano A, Czernichow P, Breant B (1998) Postnatal somatic growth and insulin contents in moderate or severe intrauterine growth retardation in the rat. Biol Neonate 73 (2):89–98
50. Zhang Y et al (2010) Effects of maternal food restriction on physical growth and neurobehavior in newborn Wistar rats. Brain Res Bull 83(1–2):1–8
51. Carney EW, Kimmel CA (2007) Interpretation of skeletal variations for human risk assessment: delayed ossification and wavy ribs. Birth Defects Res B Dev Reprod Toxicol 80(6):473–496
52. Tyl RW (2012) Commentary on the role of maternal toxicity on developmental toxicity. Birth Defects Res B Dev Reprod Toxicol 95 (3):262–266
53. Daston GP, Seed J (2007) Skeletal malformations and variations in developmental toxicity studies: interpretation issues for human risk assessment. Birth Defects Res B Dev Reprod Toxicol 80(6):421–424
54. Collins TF et al (1987) Potential reversibility of skeletal effects in rats exposed in utero to caffeine. Food Chem Toxicol 25(9):647–662
55. Marr MC et al (1992) Developmental stages of the CD (Sprague–Dawley) rat skeleton after maternal exposure to ethylene glycol. Teratology 46(2):169–181
56. Wilson GN, Howe M, Stover JM (1985) Delayed developmental sequences in rodent diabetic embryopathy. Pediatr Res 19 (12):1337–1340
57. Reuter U et al (2003) Evaluation of OECD screening tests 421 (reproduction/developmental toxicity screening test) and 422 (combined repeated dose toxicity study with the reproduction/developmental toxicity screening test). Regul Toxicol Pharmacol 38 (1):17–26
58. Takayama S et al (1995) A collaborative study in Japan on optimal treatment period and parameters for detection of male fertility disorders induced by drugs in rats. J Am Coll Toxicol 14:266–292
59. Ulbrich B, Palmer AK (1995) Detection of effects on male reproduction-a literature survey. J Am Coll Toxicol 14:293–327
60. Cooper KJ, Haynes NB, Lamming GE (1970) Effects of unrestricted feeding during oestrus on reproduction in the underfed female rat. J Reprod Fertil 22(2):293–301
61. Piacsek BE, Meites J (1967) Reinitiation of gonadotropin release in underfed rats by constant light or epinephrine. Endocrinology 81 (3):535–541
62. Sterin AB et al (1983) In vitro contractile responses of the uterus from ’restricted diet’ rats to adrenoceptor agonists. Influence of cyclo-oxygenase inhibitors. Eur J Pharmacol 90(4):411–417
63. Tropp J, Markus EJ (2001) Effects of mild food deprivation on the estrous cycle of rats. Physiol Behav 73(4):553–559
64. Goldey ES et al (1995) Developmental exposure to polychlorinated biphenyls (Aroclor 1254) reduces circulating thyroid hormone concentrations and causes hearing deficits in rats. Toxicol Appl Pharmacol 135(1):77–88
65. Zoeller RT, Crofton KM (2000) Thyroid hormone action in fetal brain development and potential for disruption by environmental chemicals. Neurotoxicology 21(6):935–945
66. Tan SW, Zoeller RT (2007) Integrating basic research on thyroid hormone action into screening and testing programs for thyroid disruptors. Crit Rev Toxicol 37(1–2):5–10
67. Bowman CJ et al (2003) Effects of in utero exposure to finasteride on androgen-dependent reproductive development in the male rat. Toxicol Sci 74(2):393–406
68. Macleod DJ et al (2010) Androgen action in the masculinization programming window and development of male reproductive organs. Int J Androl 33(2):279–287
69. McIntyre BS, Barlow NJ, Foster PM (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(2):236–249
70. Welsh M et al (2010) Critical androgen-sensitive periods of rat penis and clitoris development. Int J Androl 33(1):e144–e152
71. Goldman AS, Shapiro B, Neumann F (1976) Role of testosterone and its metabolites in the differentiation of the mammary gland in rats. Endocrinology 99(6):1490–1495
72. Topper YJ, Freeman CS (1980) Multiple hormone interactions in the developmental biology of the mammary gland. Physiol Rev 60 (4):1049–1106
73. Imperato-McGinley J et al (1986) Nipple differentiation in fetal male rats treated with an

inhibitor of the enzyme 5 alpha-reductase: definition of a selective role for dihydrotestosterone. Endocrinology 118(1):132–137

74. Imperato-McGinley J et al (1992) Comparison of the effects of the 5 alpha-reductase inhibitor finasteride and the antiandrogen flutamide on prostate and genital differentiation: dose–response studies. Endocrinology 131 (3):1149–1156
75. Wolf CJ et al (2002) Effects of prenatal testosterone propionate on the sexual development of male and female rats: a dose–response study. Toxicol Sci 65(1):71–86
76. McIntyre BS, Barlow NJ, Foster PM (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(1):62–70
77. Parker R (2012) Reproductive toxicity testing-methodology. In: Hood RD (ed) Developmental and reproductive toxicology: a practical approach, 3rd edn. Informa Healthcare, London and New York, pp 184–221
78. Gallavan RH Jr et al (1999) Interpreting the toxicologic significance of alterations in anogenital distance: potential for confounding effects of progeny body weights. Reprod Toxicol 13(5):383–390
79. Stoker TE et al (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(2):197–252
80. Clark RL (1999) Endpoints of reproductive system development. In: Daston GP, Kimmel CA (eds) An evaluation and interpretation of reproductive endpoint for human risk assessment. International Life Sciences Institute, Health and Environmental Science Institute, Washington, DC, pp 27–62
81. Ashby J, Lefevre PA (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 (1):35–47
82. Marty MS, Johnson KA, Carney EW (2003) Effect of feed restriction on Hershberger and pubertal male assay endpoints. Birth Defects Res B Dev Reprod Toxicol 68(4):363–374
83. European Commission, European Commission Regulation (EU) 2015/282 amending Annexes VIII, IX and X to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) as regards the Extended One-Generation Reproductive Toxicity Study. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2015.050.01.0001.01.ENG. Accessed 18 Nov 2015
84. ECHA (2015) European Chemicals Agency. Guidance on information requirements and chemical safety assessment—Chapter R.7a: endpoint specific guidance, version 4.1, October 2015. ISBN: 978-92-9247-411-9 Available at: http://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf. Accessed 18 Nov 2015
85. Stevens KR, Gallo MA (1982) Practical considerations in the conduct of chronic toxicity studies. In: Hayes AW (ed) Principles and methods of toxicology. Raven, New York, p 62
86. Anisimov VN et al (2012) Light-at-night-induced circadian disruption, cancer and aging. Curr Aging Sci 5(3):170–177
87. Flint AP et al (1986) The effect of day length on the duration of pregnancy and the onset of parturition in the rat. Q J Exp Physiol 71 (2):285–293
88. Gozeri E et al (2008) The effect of circadian rhythm changes on fetal and placental development (experimental study). Neuro Endocrinol Lett 29(1):87–90
89. Varcoe TJ et al (2011) Chronic phase shifts of the photoperiod throughout pregnancy programs glucose intolerance and insulin resistance in the rat. PLoS One 6(4):e18504
90. Cooper RL et al (2006) A tiered approach to life stages testing for agricultural chemical safety assessment. Crit Rev Toxicol 36 (1):69–98
91. Chapin RE, Sloane RA, Haseman JK (1997) The relationships among reproductive endpoints in Swiss mice, using the reproductive assessment by Continuous Breeding database. Fundam Appl Toxicol 38(2):129–142
92. Gray L et al (1989) Correlation of sperm and endocrine measures with reproductive success in rodents. In: Burger EJ (ed) Progress in clinical and biological research. Liss, New York, pp 193–209
93. Mangelsdorf I, Buschmann J, Orthen B (2003) Some aspects relating to the evaluation of the effects of chemicals on male fertility. Regul Toxicol Pharmacol 37(3):356–369
94. Hirshfield AN (1987) Histological assessment of follicular development and its applicability

to risk assessment. Reprod Toxicol 1(1):71–79

95. Regan KS et al (2005) STP position paper: ovarian follicular counting in the assessment of rodent reproductive toxicity. Toxicol Pathol 33(3):409–412
96. OECD (2011) Guidance document on the current implementation of internal triggers in the extended one generation reproductive toxicity study in the United States and Canada, Series on Testing and Assessment, No. 117, ENV/JM/MONO(2011)21, Paris
97. Piersma AH et al (2011) Combined retrospective analysis of 498 rat multi-generation reproductive toxicity studies: on the impact of parameters related to F1 mating and F2 offspring. Reprod Toxicol 31(4):392–401
98. Fegert I et al (2012) Feasibility of the Extended One-Generation Reproductive Toxicity Study (OECD 443). Reprod Toxicol 34(3):331–339
99. Tonk EC et al (2013) Developmental immunotoxicity of ethanol in an Extended One-Generation Reproductive Toxicity Study. Arch Toxicol 87(2):323–335
100. Marty MS et al (2009) Inter-laboratory control data for reproductive endpoints required in the OPPTS 870.3800/OECD 416 reproduction and fertility test. Birth Defects Res B Dev Reprod Toxicol 86(6):470–489
101. Stump DG et al (2014) Key lessons from performance of the U.S. EPA Endocrine Disruptor Screening Program (EDSP) Tier 1 male and female pubertal assays. Birth Defects Res B Dev Reprod Toxicol 101(1):43–62
102. Mazzocchi G et al (1980) An ultrastructural stereologic study of the effects of angiotensin II on the zona glomerulosa of rat adrenal cortex. Acta Endocrinol (Copenh) 95(4):523–527
103. Otis M et al (2007) The growth-promoting effects of angiotensin II in adrenal glomerulosa cells: an interactive tale. Mol Cell Endocrinol 273(1–2):1–5
104. United Nations (2007) Globally harmonized system of classification and labelling of chemicals (GHS). Part 3 Health Hazards

Methods in Pharmacology and Toxicology (2017) 185–203
DOI 10.1007/7653_2014_13
© Springer Science+Business Media New York 2014
Published online: 02 October 2014

The Influence of Environmental Contaminants and Lifestyle on Testicular Damage and Male Fertility

Sellappan Selvaraju, Meritxell Jodar, and Stephen A. Krawetz

Abstract

Environmental contaminants and lifestyle play a major role in influencing human fertility. Most substances toxic to male fertility target functions of specific testicular cells or act at various levels of the hypothalamo-hypophyseal-testicular axis. The bio-accumulation of toxic compounds is often exacerbated by lifestyle choices and workplace environment. Inadvertently this impairs spermatogenesis, fertility, and survival of the offspring by affecting the genetic constitution of the spermatozoa. Such changes can be transmitted to future generations and may affect their fertility and health. Testicular insult can be assessed by sperm functional assays and by measuring biomarkers in tissues and body fluids including spermatozoa. However, histopathology remains the standard and is generally used to confirm the results of these other assays. The need of sensitive, reliable, robust, and easily measured biomarkers that are able to detect testicular injury before it becomes a chronic and irreversible effect is evident. The noninvasive approach of identifying sperm molecular signatures as biomarkers is promising and may emerge as a valuable diagnostic tool in the near future.

Key words Male infertility, Endocrine disruptors, Heavy metals, Phthalates, Pesticides, Lifestyle, Radiations, Testicular damage, Biomarkers, Transgenerational effect

1 Introduction

Infertility has almost doubled from 8 to 15 % over the past two decades. Though the reasons are many, in the industrialized countries, environmental contaminants [1] as well as lifestyle (reviewed in [2]) affect the fertility of the couple. As part of industrialization and modernization we have observed an increase in heavy metal toxicity, e.g., lead, cadmium, chromium, and associated infertility. In addition, the population explosion is leading to an increased demand for food which has pressured the agricultural industry to increase productivity to which many have turned to fertilizers and pesticides. Such practices result in the accumulation of toxic substances in the environment. These environmental contaminants have many adverse effects on various systems including the reproductive fitness that affects the birth and life of the child.

Toxic substances may modify the genome, transcripts, and proteins expressed in the testis with their results specifically felt in

spermatozoa (reviewed in [3]). For example, paternal habits such as smoking can yield germ line mutations that accumulate and are transmitted to the offspring [4]. In a similar manner, environmental factors can induce epigenetic germ line changes that are transmitted to the next generation [5]. On one hand, transgenerational epigenetic inheritance may increase the competitiveness of the offspring as it can provide a mechanism that allows them to adapt to new environments to which the parents were exposed. This is exemplified in the case of the transmission of the traumatic olfactory response to the offspring as an innate defense mechanism against predators [6]. On the other hand, changes in chromatin state such as methylation or structure could transmit a predisposition to disease (reviewed in [7]). One must also be aware and consider the potential epigenetic function of spermatozoal RNAs [8] that perhaps influences transgenerational inheritance (reviewed in [9]).

Spermatogenesis is a well-orchestrated hormonally regulated process within the testicular compartment leading to the production of mature spermatozoa from a spermatogonium. This is a tightly regulated process mediated by a positive/negative feedback mechanism of hormones secreted from the hypothalamo-hypophyseal-gonadal axis. As summarized in Fig. 1, the hypothalamus releases gonadotropin-releasing hormone (GnRH), which signals the hypophysis to release gonadotropins. These include, Follicle-Stimulating Hormone (FSH) and Luteinizing Hormone (LH) where LH signals the Leydig cells to produce testosterone. In turn, high levels of testosterone provide negative feedback to the hypothalamus and hypophysis leading to the regulated release of GnRH and gonadotropins. FSH acts in a synergistic manner with testosterone on Sertoli cells stimulating the secretion of inhibin and the Androgen-Binding Protein (ABP). The binding of testosterone to ABP ensures the appropriate local concentration of testosterone guaranteeing that spermatogenesis completes. Paradoxically, spermatogenesis requires FSH and a high level of intracellular testosterone but the germinal cells are void of FSH and testosterone receptors. This is circumvented by Sertoli cell tight junctions (Fig. 1) where Sertoli cells release inhibin to provide negative feedback for FSH. Endocrine disruptors, like phthalates (Table 1), can upset the hypothalamo-hypophyseal axis as it suppresses Leydig cell as well as Sertoli cell support.

The transgenerational influence of a father to the offspring's phenotype as a function of environmental exposure whether it be changes in nutritional status or exposure to endocrine disruptors is the subject of intense investigation ([5], reviewed in [10, 11]). For example, nutrient intake and food composition may affect the offspring's development and phenotype. Although these effects can be diverse, this is best illustrated by caloric restriction of the male prior to mating altering the metabolism of the offspring [12]. These include altered lipid and cholesterol biosynthesis of offspring

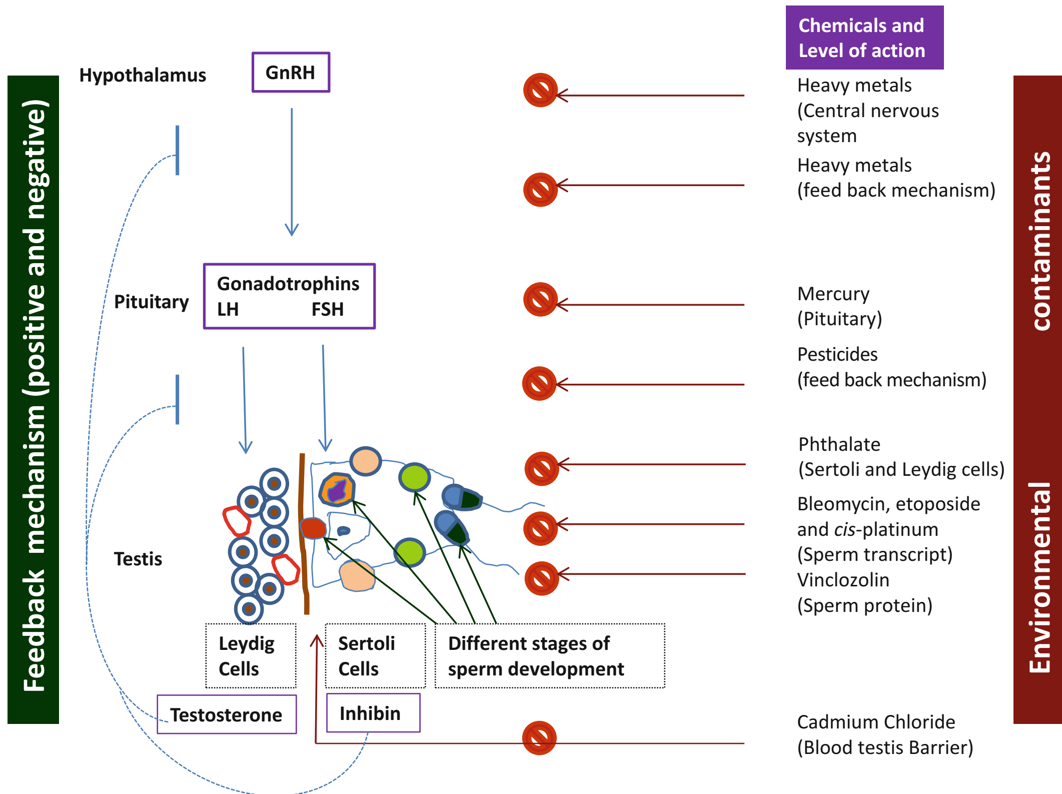

Fig. 1 Influence of environmental contaminants and their probable level of action on the hypothalamo-hypophyseal-testicular system. Endocrine disruptors affect hormonal synthesis and secretion by acting either at hypothalamus, pituitary, and testicular levels (e.g., heavy metals) or by positively or negatively affecting hormonal feedback (e.g., pesticides). Certain chemicals affect the integrity of the Leydig and Sertoli cells (e.g., phthalates) and blood-testes barrier (e.g., cadmium chloride) leading to changes in testicular architecture and composition. Some chemicals alter the levels of sperm transcripts (e.g., bleomycin, etoposide, and *cis*-platinum) and proteins (e.g., vinclozolin) and may affect fertility across generations. Understanding the pathway of action can help to elucidate the role of toxicants on male fertility and integrity

derived from male mice raised on a low-protein diet [13] or reduced glucose tolerance of offspring derived from males raised on a high-fat diet [14]. Similar effects are observed in humans where an increased risk for diabetes and cardiovascular diseases is linked to overweight fathers and grandfathers [15]. The outcome can reflect the time of exposure. For example, paternal exposure to DiButyl Phthalate (DBP) alters the sex ratio of the offspring, delays female sexual maturation, and degrades F1 sperm quality [16]. Further, developing embryos exposed to endocrine disrupters can transgenerationally affect spermatogenesis [5]. Understanding the role of epigenetics as a function of nutrition as well as endocrine disruptors may provide insight into the mechanism of action of these substances providing markers for assessing fertility and rational treatment options. Accordingly, the effect of various environmental

Table 1
Potential toxic substances affecting various cells in testis

Chemical	Target	Effect on the germ cell/ sperm function/fertility	References
2-Methoxyethanol,	Pachytene spermatocyte	Death of the pachytene spermatocyte	[135, 136]
Methoxyacetic acid (MAA)	Pachytene spermatocyte	Disrupts pachytene spermatocytes	[116]
Dinitrobenzene	Sertoli cell	Death of the pachytene spermatocyte	[116, 137, 138]
Ethane dimethane sulfonate	Leydig cell	Death of the pachytene spermatocyte Death of the spermatid	[139]
2,5-Hexanedione	Sertoli cell	Death of the pachytene spermatocyte Round and elongating spermatid-death	[140]
Mono-2-ethylhexyl phthalate (MEHP)—	Sertoli cell	Development of the germ cell lineage in humans	[141]
Cadmium chloride	Blood-testis barrier		[116]
Chromium(VI)	Sertoli cells Testicular atrophy	Disruption in germ cell arrangement	[51, 142]
Molybdenum	Testicular architecture	Sperm concentration and morphology	[45]
Mercury	Hypothalamus and pituitary	Impair GnRH and gonadotropins release	[33, 34]
Glyphosate	Leydig cells	Spermatogenesis and testosterone production	[77]
Reetha saponins and quinine hydrochloride (Azadirachta Indica)	Spermatozoa	Spermicide	[143]
Smoking	Leydig cells and Sertoli cells	Spermatogenesis	[102, 103]
Sulfur mustard gas	Germ cells	Spermatogenesis, azoospermia	[144]
Methyl isocyanate	Fetotoxic	Affect testicular function	[145]
Thallium	Sertoli cells, germ cells	Sperm cell development and function	[58]
Alcohol(Ethanol)	Male fertility	Hypogonadism, testicular atrophy, hypothalamo-hypophyseal-gonadal axis, Leydig and Sertoli cell function, feminization	[146]
Pesticides	HPG axis	Disruption in endocrine function	[147]

contaminants and associated testicular toxic effects are discussed with respect to novel tests that are on the horizon to assess the testicular damage.

2 Chemicals Affecting Male Fertility

2.1 Phthalates

Phthalates are widely used in many consumer products including personal care, food wrappers, plastic containers, paints, floorings, wall coverings, and medical equipment (reviewed in [17]). This results in continuous and ubiquitous exposure. Phthalates, mainly diester phthalates (DEHP) and DBP, are used in personal care products and are known to have weak estrogenic [18] as well as antiandrogenic activity but without androgen receptor binding. They primarily interfere with key genes involved in cholesterol transport and testosterone biosynthesis (reviewed in [19]). DEHP is metabolized into primary (mono-2-ethylhexyl phthalate, MEHP) and secondary metabolites and then excreted into the urine. The percentage of the primary metabolite, MEHP, excreted through urine inversely correlates with reproductive steroid hormone levels [20–22].

The effect of phthalates on male reproduction is not well understood. Phthalates primarily act as endocrine disruptors and in males their effect has been termed the phthalates syndrome. Metabolites such as monoethyl phthalate (MEP), MEHP, monobenzyl phthalate (MBzP), monobutyl phthalate (MBP), and MEHP in urine are now being evaluated as biomarkers to assess phthalate exposure (reviewed in [17, 20, 23]). In utero phthalate exposure disrupts the formation of the male reproductive tract presenting a host of affects. These include reduced anogenital distance, cryptorchidism, reduction in fetal testosterone, abnormally located Leydig cells, abnormal seminiferous cord formation (reviewed in [24]), and reduced antioxidant enzyme levels in testes [25]. Higher urinary concentrations of phthalate metabolites have been observed in infertile couples undergoing assisted reproductive techniques suggesting that the phthalates could affect fertility [26]. Several studies have also shown the negative influence of phthalate exposure in adults on seminal parameters (reviewed in [17, 27]) and sperm DNA fragmentation [28].

2.2 Heavy Metals

We are continuously exposed to heavy metals from various sources such as contaminated air, water, soil, food, herbal medicines, and consumer products (reviewed in [29]). Some of the heavy metals reported to affect human health and male fertility are cadmium (Cd), lead (Pb), molybdenum (Mb), chromium (Cr), mercury (Hg), and arsenic (As) (reviewed in [30]). Industrialization has been associated with an increased frequency of infertility mainly

due to the release of these toxic heavy metals and their resulting environmental contamination. They have been shown to affect testicular function, semen quality, and fertility in both humans [31] and animals [32] and can be considered a widespread cause for concern. Heavy metals may affect neurotransmitter function and impair the pulsatile release of GnRH from the hypothalamus [33]. Apart from their action on the hypothalamus, certain heavy metals, e.g., mercury, are stored in the pituitary and affect the release of gonadotropins [34]. Heavy metals generate reactive oxygen species and resulting oxidative stress leads to cell injury (reviewed in [30]). Heavy metals also affect various endocrine glands, including adrenal, and thyroid leading to impaired reproductive function. These metals have a long half-life ranging from 10 to 35 years (reviewed in [35]), and as such can have a cumulative effect.

Cadmium toxicity usually occurs in areas with ground water contaminated mainly from agriculture that relies on the use of phosphate fertilizers. Tobacco is a prime accumulator and smoking further increases the body burden. Cadmium toxicity induces seminiferous tubule necrosis and interstitial testis edema ([36], reviewed in [37]). Effects on spermatozoal motility, membrane integrity as well as mitochondrial membrane integrity and DNA damage impairing sperm function have been documented [32, 38, 39]. While cadmium toxicity may be more of a concern [32] seminal plasma Pb at a concentration >10 μg/l may have significant effect on male reproduction [39]. This includes a reduction in sperm concentration [40], motility [41], morphology [42], and impaired chromatin condensation [43].

Molybdenum is reported to be toxic in certain areas in close proximity to industrial effluents and chemical reagents [44, 45]. Molybdenum is an endocrine disruptor (reviewed in [46]) and its concentration is inversely associated with circulating testosterone [47]. Males appear more prone to molybdenum toxicity as compared to females [48]. Molybdenum toxicity may lead to testicular degeneration [44], reduced levels of testicular antioxidant enzymes [49], reduced sperm concentration, and abnormal morphology [45]. Fertilization with spermatozoa exposed to toxic levels of molybdenum exhibits male-mediated embryo toxicity [44].

Apart from the above, chromium is also implicated in male infertility. Chromium is found in all natural substances including rocks, plants, soils, gases, as well as industrial waste. This element potentiates insulin action affecting glucose, protein, and fat metabolism. It is also known to induce testicular atrophy and reduce sperm concentration and motility [50, 51], though earlier studies reported no adverse effect on semen parameters and spermatozoal function [52]. Seminal plasma chromium has a significant toxic effect on sperm production and motility [53, 54]. This is partly attributed to its antagonistic effect on the level of seminal plasma

zinc [51] as zinc is known for its role as an antioxidant. Zinc deficiency leads to higher levels of oxidative damage to the testicular cells and associated infertility ([55], reviewed in [56]).

Thallium is used in a variety of consumer products including cosmetic jewelry, optical lenses, the manufacture of cement, and semiconductors and is a by-product of the combustion of coal that can be used as a rodenticide. Thallium salts are also used to treat syphilis, gonorrhea, and ringworm as they can induce a potent systemic toxicity by inhibiting glycolysis, Kreb's cycle, and oxidative phosphorylation. In males, chronic thallium toxicity affects Sertoli cell and germ cell function [57, 58] but is also ovotoxic crossing the placental barrier leading to fetal abnormalities and/or fetal death [59].

2.3 Chemotherapy and Radiation

Spermatogonia and rapidly dividing differentiated mitotic cells are the most sensitive to chemotherapy as compared to spermatocytes undergoing meiotic division. In comparison, the spermatogonial stem cells are of intermediate sensitivity [60]. Chemotherapy results in either temporary or permanent azoospermia (the absence of sperm) and this effect depends on the dose and duration of the treatment (reviewed in [61, 62]). While high dose chemotherapy is associated with irreversible germinal epithelial failure (reviewed in [63]), Leydig cells appear to be resistant when patients are treated by chemotherapy [64]. Nevertheless, the severity of chemotherapy on Leydig cell function may depend on other factors like age and dose of the drugs [64–66]. Chemotherapy-induced gonadal toxicity is immediate but reversible in the adult male as compared to prepubertal males (reviewed in [62]). For example, treatment of testicular cancer with cisplatin and carboplatin regimens leads to a loss or reduced spermatozoa concentration in semen of most men, with a recovery to ~80 % normospermia by 5 years (reviewed in [63]).

X-ray and electromagnetic ionizing radiation can affect semen quality concentration, motility, morphology, and DNA integrity [67] in the patient undergoing treatment [68] as well as in the occupational setting [69]. Spermatogonia are most susceptible to ionizing radiation compared to spermatocytes and spermatids. Leydig cell function is also affected in tumor patients treated with high-dose radiation [70]. The severity of impact depends on various factors of dose, duration, location of therapy, age of the individual, systemic effect of the disease, endocrine disturbance, and autoimmune disorders (reviewed in [71, 72]).

2.4 Pesticides and Herbicides

Several pesticides and herbicides possess estrogenic and androgenic activities that essentially impede reproductive function. Typically they interact with their receptor to block the induction of expression (reviewed in [73]). This typically has two effects, delayed puberty and infertility [74, 75]. For example, glyphosate-based pesticides are in routine use worldwide often used in conjunction with genetically modified plants grown for food and feed (reviewed in [76]).

They are major pollutants of rivers and surface water. Even at 1 ppm they damage Leydig cells and testosterone production is compromised [77]. At 36 ppm Sertoli cells are subject to oxidative stress that defeats the antioxidant mechanisms and dysregulates cell signaling. This leads to cell death [78], DNA damage, birth defects, liver dysfunction, and cancer in variety of cell types ([77], reviewed in [76]). Even with the above plethora of negative impacts, the effect of glyphosate on reproductive function remains controversial (reviewed in [79]). Similarly, alachlor [2-chloro-2',6'-diethyl-*N*-(methoxymethyl)-acetanilide], a herbicide commonly used in agriculture is also an endocrine disruptor [80]. Men exposed to alachlor are known to have lowered sperm motility [1] reflective of apoptosis affecting sperm mitochondrial membrane integrity and DNA integrity [81]. Germ line epigenetic changes in imprinted genes expressed even in the F3 offspring have been observed when pregnant dams were exposed to the agricultural pesticide at the time of sex determination. Similarly, exposure of the pregnant dam to pesticides can perturb the F1 sperm epigenome and that of the Sertoli cell that by the F3 resolves as male infertility [11].

2.5 Lifestyle, Nutrition, Alcohol, and Smoking

Nutrition regulates fertility at various levels, i.e., reproductive tract development, gametogenesis, fertilization, and embryo development that phenotypically resolves in the offspring. Foods rich in anabolic steroids and xenobiotics may lower male fertility. For example, beef raised on synthetic estrogen and anabolic steroids or from infertile animals treated with steroids may have a harmful effect on male fertility [82]. Mammalian spermatozoa contain high levels of membrane-associated polyunsaturated fatty acids [83], to maintain membrane fluidity and flexibility associated with sperm motility. Diets with different sources of fat modify the lipid content of the sperm head, midpiece, and tail membranes ultimately influencing semen quality [84]. On one hand, consumption of dairy foods rich in fat is known to reduce sperm concentration and progressive forward sperm motility [85]. On the other hand, nutraceuticals rich in *n*-3 fatty acids [86–88] and also those with certain ratios of *n*-3/*n*-6 fatty acids [89] are being considered as a means to improve male fertility (reviewed in [90]). However their effect and efficacy remains to be resolved.

Smoking and alcoholism are two important lifestyles variables that affect the fertilizing ability of the gametes and embryo survival. Chronic alcohol abuse can affect Sertoli cell mRNAs and protein expression compromising function [91]. In turn this may lead to testicular atrophy and reduced testosterone production and metabolism [92–94]. This reduces sperm concentration and impairs morphology [92, 95] presenting as subfertility in humans. Paternal alcohol abuse impairs enzyme function and may alter spermatozoal RNAs with its initial effect felt during fetal development [94, 96, 97]. A transgenerational epigenetic effect is also likely.

It has been estimated that 30–35 % of the men and women of reproductive age are smokers in the United States [98]. Tobacco smoking contains heavy metals like cadmium, lead, and thallium, resulting in heavy metal toxicity [99, 100]. Though the effect of smoking on male fertility was not convincingly demonstrated, various studies suggest a possible negative effect of smoking on sperm parameters [101] and fertility [102, 103]. Smoking induces spermatozoal DNA damage [104] that may induce embryo damage [105]. Current efforts are now focused towards the changes in spermatozoal mRNAs associated with smoking leading to infertility [106] including spermatozoal miRNAs [107]. Such an effect may be of epigenetic concern and could influence the phenotype of the offspring [107, 108].

3 Methods

3.1 Assessment of Testicular Toxicity

While the studies in animal models are well controlled the effects of any toxic compound are typically assessed in humans from an epidemiologic and retrospective perspective. Several noninvasive and invasive techniques are in current use to assess the effect of various insults on testes and male reproductive function (Table 2). Noninvasive techniques including mating ratios, pregnancy ratios, pre- and post-implantation loss, litter size, number of live and dead pups, sex ratio, and offspring survival are tested after the exposure. While the noninvasive methodologies can be very informative, the comparative invasive techniques are primarily reserved for use in animal models, as they can yield the mechanism that produces the infertility.

The study of testicular toxicity of different compounds in animals allows one to determine the characteristics of exposure such as dose–response, time course, stage specificity, target cells, and mode of action. Several techniques are useful to assess outcomes including changes in the weight of reproductive organs such as testis, epididymis, and prostate. After toxic exposure, a variation in the weight of male gonads could be indicative of a potential adverse effect. On one hand, an increase in organ weight following treatment would suggest an increase in fluid content in the tissue due to, e.g., inflammation after exposure to the insecticide cypermethrin [109]. On the other hand, after cadmium exposure a decrease in the weight of testis or epididymis is observed as spermatogenesis is blocked decreasing the production of spermatozoa [110]. Another useful outcome is enzyme activity, including testicular alkaline phosphatase, acid phosphatase, lactate dehydrogenase, glucose-6-phosphate dehydrogenase, and various steroidogenic enzymes. The activities of these enzymes echo various states of spermatogenesis, including oxidative stress and testicular steroidogenesis.

Histopathology is the gold standard of testicular toxicity as it provides an exact description of the areas affected. This facilitates

Table 2
Various approaches for evaluation of testicular toxicity

Markers/methods for assessing testicular toxicity	Remarks	Reference
Inhibin B (serum or plasma)	Widely employed[a]	[148]
Testosterone	[b]	[149]
Follicle-stimulating hormone	[c]	[150]
Luteinizing hormone	[c]	[150]
Sperm transcriptome/epigenome/ metabolome/testis transcriptome	[b]	[129, 151]
Tissue-specific circulatory miRNAs	[c]	[122]
Semen evaluation tests Sperm morphology, concentration, viability, sperm functional membrane integrity tests	[b]	[32, 149]
Testicular enzymes Hyaluronidase, sorbitol dehydrogenase, lactate dehydrogenase isoenzyme-X	[b]	[152, 153]
Serum marker levels		
Germ cell markers in Interstitial fluid Phosphatidylethanolamine-binding protein Androgen-regulated protein 2	Diagnostic value. Presence of germ cell marker in blood may be due to germ cell damage or blood-testes barrier damage	[154]

[a]Widely employed, [b]Moderately employed and [C]Less employed based on the published literature on this aspect

identifying the primary target as well as possible pathways involved during the course of exposure to a toxic substance(s) (reviewed in [111]). One often overlooked consideration is the potential for reversion of toxicity if the spermatogonial stem cell has not been modified since spermatogenesis is a continuous cycle the length of time of which is governed by the species [112]. In comparison, the effect of reversal on Leydig or Sertoli and other support cells is less obvious. Nevertheless, the notion that the effects of some toxins could be reversed offers an interesting possibility towards developing reversible male contraceptives.

Though the research is being carried out in laboratory animals to extrapolate the effect of environmental factors on spermatogenesis in human, its interpretation warrants caution as certain toxins can have profound effect on the human reproduction as compared to animals. This is best exemplified by ionizing radiation that has a threefold increase in effect on human spermatogenesis as compared to mouse [113].

3.2 Biomarkers of Testicular Damage and Male Fertility in Humans

Exposure to toxic compounds when paired with our unhealthy habits is not as well controlled as in animal models. Although exposure to these compounds can be assessed by questionnaires, new approaches allow the precise evaluation of the exposure using urine or blood samples. For example, the presence of some primary or secondary metabolites of phthalates in urine or the presence of cotinine, a metabolite of nicotine, in serum can be used to assess the exposure reflective of half-life. Consider dichlorodiphenylchloroethane (DDE) that has a half-life of approximately 6 years as compared to parent chemical DDT which has a half-life of 4 years. In such a case the proportion of the metabolite to parent compound in the body fluids can be used to calculate the length of the exposure (reviewed in [114]).

The analysis of hormones in plasma is useful to detect any derangement of endocrine signaling fertility related pathways typified by the effects of endocrine disruptors and thiols. Some like Bisphenol A (BPA), dichlorodiphenyltrichloroethane (DDT), and PBC plastic imitate estrogens whereas others such as the fungicide vinclozolin act through the androgen receptor (reviewed in [24]). Other compounds such as mercury mainly affect the hypothalamus and pituitary impairing the release of GnRH and gonadotropins. The mimic action of these exogenous components deregulates endocrine signaling altering the levels of endogenous hormones typically detected in plasma. Currently their effects are assessed as a function of seminal parameters and circulating plasma levels of testosterone, LH, and FSH. These parameters present several challenges that are compounded by the high intra-individual variability in seminal parameters. Further, only severe and potentially irreversible injuries are usually detected in the clinic by significantly altered hormone levels like testosterone, FSH, LH, and Inhibin B [115]. Other small molecules in plasma that could assess testicular damage include the proteins ADAM3, Calpastatin, DAZL, FABP9, and VASA that are specific to the seminiferous tubules bathed in interstitial fluid. However, unless the blood-testis barrier (BTB) is compromised it is impossible to detect these proteins [116]. Accordingly, the use of such biomarkers has limited potential to infer testicular response and is best suited to compounds that compromise the BTB like cadmium. Although the gold standard, assessing testicular damage using histopathology is generally difficult to clinically deploy. There is reluctance to biopsy since it typically yields small samples which may be not representative of all cell types present in testis.

Several studies have correlated sperm parameters with different environmental factors but have yielded conflicting results (reviewed in [117]). This was likely due to multiple sources of variation including age [118], geographic region [119], smoking status (reviewed in [120]), and period of abstinence before collection [121]. Together these results suggest that evaluation of semen quality alone is not sufficient to establish the effects of environmental factors on male fertility.

This emphasizes the need to develop new surrogates of exposure that are easily accessible, reliable, and sensitive so as to provide an early warning measure before irreversible damage occurs. These new biomarkers can be divided into three classes, small molecules which include hormones and single compounds, proteins and nucleic acids, both RNA and DNA that are found in body fluids including saliva, blood, and semen. The potential use of circulating miRNAs specific from testes that could assess testicular damage is just being considered but their utility is uncertain [115, 122].

It has been suggested that some environmental toxicants can induce DNA damage in the sperm nuclear as well as the mitochondrial genomes ([123], reviewed in [37, 124]). Several tests are now available to assess spermatozoal DNA damage. This is particularly useful in cases of chemotherapy where damage is reflected by a direct increase in sperm aneuploidies that can be identified by FISH [125]. DNA methylation and sperm chromatin structure analysis may also provide a clue of epigenetic changes induced by environmental factors that could be transmitted to the offspring.

Spermatozoa retain a complex population of RNAs [9, 126] that could be used to assess past events, like exposures and their effect on spermatogenesis. Their use as markers of fertility status has been suggested [127, 128]. Some evidence in animal models suggests that the spermatozoal RNA profile is a better biomarker to assess the impact of environmental factor on male fertility than hormones or semen parameters tests [129].

An extended catalog of sperm proteins is available (reviewed in [130]) that could also be used as biomarkers of testicular damage. For example, exposure to endocrine disrupting chemicals affects several proteins including malate dehydrogenase 2, aldehyde dehydrogenase, and A-kinase anchor protein 4 that are known to play a crucial role in sperm fertilizing potential [131]. A decrease in protamination along with an increase in histones retained in the sperm head has been observed in rats given bleomycin, etoposide, and *cis*-platinum [132]. Testicular toxicity is also observed upon treatment with diethyl stilboesterol by downregulating thioredoxin-like-1 protein that leads to a cascade of events associated with oxidative stress [133] impacting fertility. Like RNA, proteomic approaches may be able to differentiate between subtypes of infertile patients [134] which together may introduce a new class of biomarkers of male toxicity.

4 Conclusion

Environmental exposures are ubiquitous arising from many sources having both a direct and indirect impact through, e.g., contaminated feed, water, and air. For example, endocrine disrupters can induce testicular damage by dysregulating hormone synthesis and

secretion that also impacts antioxidant status, impacting testes integrity, sperm structure, and composition. Various diagnostic markers have been explored to identify the nature and extent of exposure to toxic substances and their consequences on male fertility. For the moment invasive histopathology techniques provide the best single characterization of the toxic response. This has limited their routine clinical use highlighting the need for new biomarkers to assess testicular damage before its chronification. The alteration of different spermatozoal molecular signatures (DNA, RNA, and proteins) resulting from exposure to toxic chemicals appears to be on the horizon. The in-depth analysis of the molecular changes undertaken by sperm will also permit us to clarify the mechanisms of transgenerational epigenetic inheritance and its long-term impact. Irrespective, and until this is understood, providing a healthy environment and maintaining a similar lifestyle should be prescribed as a means to maintain fertility in the present as well preserve the health of future generations.

Acknowledgements

We would also like to thank many colleagues who criticaly reviewed this manuscript. apologize to others in the field that for space limitations we could not directly reference their contributions to the field.

Funding

S.S. is supported by a Cutting-edge Research Enhancement and Scientific Training Award, Department of Biotechnology, Government of India. Support from the Charlotte B. Failing Professorship to SAK is gratefully acknowledged.

References

1. Swan SH, Kruse RL, Liu F et al (2003) Semen quality in relation to biomarkers of pesticide exposure. Environ Health Perspect 111 (12):1478–1484
2. Sharma R, Biedenharn KR, Fedor JM et al (2013) Lifestyle factors and reproductive health: taking control of your fertility. Reprod Biol Endocrinol 11:66
3. Campion S, Catlin N, Heger N et al (2012) Male reprotoxicity and endocrine disruption. EXS 101:315–360
4. Linschooten JO, Verhofstad N, Gutzkow K et al (2013) Paternal lifestyle as a potential source of germline mutations transmitted to offspring. FASEB J 27(7):2873–2879
5. Anway MD, Cupp AS, Uzumcu M et al (2005) Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science 308(5727):1466–1469
6. Dias BG, Ressler KJ (2014) Parental olfactory experience influences behavior and neural structure in subsequent generations. Nat Neurosci 17(1):89–96
7. Rando OJ (2012) Daddy issues: paternal effects on phenotype. Cell 151(4):702–708
8. Rassoulzadegan M, Grandjean V, Gounon P et al (2006) RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature 441(7092):469–474
9. Jodar M, Selvaraju S, Sendler E et al (2013) The presence, role and clinical use of spermatozoal RNAs. Hum Reprod Update 19 (6):604–624

10. Skinner MK (2007) Endocrine disruptors and epigenetic transgenerational disease etiology. Pediatr Res 61(5 Pt 2):48R–50R
11. Guerrero-Bosagna C, Savenkova M, Haque MM et al (2013) Environmentally induced epigenetic transgenerational inheritance of altered Sertoli cell transcriptome and epigenome: molecular etiology of male infertility. PLoS One 8(3):e59922
12. Anderson LM, Riffle L, Wilson R et al (2006) Preconceptional fasting of fathers alters serum glucose in offspring of mice. Nutrition 22 (3):327–331
13. Carone BR, Fauquier L, Habib N et al (2010) Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. Cell 143 (7):1084–1096
14. Ng SF, Lin RC, Laybutt DR et al (2010) Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. Nature 467(7318):963–966
15. Kaati G, Bygren LO, Edvinsson S (2002) Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. Eur J Hum Genet 10(11):682–688
16. Dobrzynska MM, Tyrkiel EJ, Pachocki KA (2011) Developmental toxicity in mice following paternal exposure to Di-N-butylphthalate (DBP). Biomed Environ Sci 24 (5):569–578
17. Hauser R (2008) Urinary phthalate metabolites and semen quality: a review of a potential biomarker of susceptibility. Int J Androl 31(2):112–117
18. Harris CA, Henttu P, Parker MG et al (1997) The estrogenic activity of phthalate esters in vitro. Environ Health Perspect 105 (8):802–811
19. Fisher JS (2004) Environmental anti-androgens and male reproductive health: focus on phthalates and testicular dysgenesis syndrome. Reproduction 127(3):305–315
20. Meeker JD, Calafat AM, Hauser R (2009) Urinary metabolites of di(2-ethylhexyl) phthalate are associated with decreased steroid hormone levels in adult men. J Androl 30(3):287–297
21. Pan G, Hanaoka T, Yoshimura M et al (2006) Decreased serum free testosterone in workers exposed to high levels of di-n-butyl phthalate (DBP) and di-2-ethylhexyl phthalate (DEHP): a cross-sectional study in China. Environ Health Perspect 114 (11):1643–1648
22. Joensen UN, Frederiksen H, Jensen MB et al (2012) Phthalate excretion pattern and testicular function: a study of 881 healthy Danish men. Environ Health Perspect 120 (10):1397–1403
23. Jonsson BA, Richthoff J, Rylander L et al (2005) Urinary phthalate metabolites and biomarkers of reproductive function in young men. Epidemiology 16(4):487–493
24. Balabanic D, Rupnik M, Klemencic AK (2011) Negative impact of endocrine-disrupting compounds on human reproductive health. Reprod Fertil Dev 23(3):403–416
25. Zhou D, Wang H, Zhang J et al (2010) Di-n-butyl phthalate (DBP) exposure induces oxidative damage in testes of adult rats. Syst Biol Reprod Med 56(6):413–419
26. Tranfo G, Caporossi L, Paci E et al (2012) Urinary phthalate monoesters concentration in couples with infertility problems. Toxicol Lett 213(1):15–20
27. Wirth JJ, Rossano MG, Potter R et al (2008) A pilot study associating urinary concentrations of phthalate metabolites and semen quality. Syst Biol Reprod Med 54(3):143–154
28. Hauser R, Meeker JD, Singh NP et al (2007) DNA damage in human sperm is related to urinary levels of phthalate monoester and oxidative metabolites. Hum Reprod 22 (3):688–695
29. Ernst E (2002) Toxic heavy metals and undeclared drugs in Asian herbal medicines. Trends Pharmacol Sci 23(3):136–139
30. Valko M, Morris H, Cronin MT (2005) Metals, toxicity and oxidative stress. Curr Med Chem 12(10):1161–1208
31. Sengupta M, Deb I, Sharma GD et al (2013) Human sperm and other seminal constituents in male infertile patients from arsenic and cadmium rich areas of Southern Assam. Syst Biol Reprod Med 59(4):199–209
32. Selvaraju S, Nandi S, Gupta PS et al (2011) Effects of heavy metals and pesticides on buffalo (Bubalus bubalis) spermatozoa functions in vitro. Reprod Domest Anim 46 (5):807–813
33. Lakshmana MK, Desiraju T, Raju TR (1993) Mercuric chloride-induced alterations of levels of noradrenaline, dopamine, serotonin and acetylcholine esterase activity in different regions of rat brain during postnatal development. Arch Toxicol 67(6):422–427
34. Danscher G, Horsted-Bindslev P, Rungby J (1990) Traces of mercury in organs from primates with amalgam fillings. Exp Mol Pathol 52(3):291–299

35. Godt J, Scheidig F, Grosse-Siestrup C et al (2006) The toxicity of cadmium and resulting hazards for human health. J Occup Med Toxicol 1:22
36. Martynowicz H, Skoczynska A, Karczmarek-Wdowiak B et al (2005) Effects of cadmium on testis function. Med Pr 56(2):167–174
37. Wirth JJ, Mijal RS (2010) Adverse effects of low level heavy metal exposure on male reproductive function. Syst Biol Reprod Med 56 (2):147–167
38. Arabi M, Mohammadpour AA (2006) Adverse effects of cadmium on bull spermatozoa. Vet Res Commun 30(8):943–951
39. Xu DX, Shen HM, Zhu QX et al (2003) The associations among semen quality, oxidative DNA damage in human spermatozoa and concentrations of cadmium, lead and selenium in seminal plasma. Mutat Res 534 (1–2):155–163
40. Wu HM, Lin-Tan DT, Wang ML et al (2012) Lead level in seminal plasma may affect semen quality for men without occupational exposure to lead. Reprod Biol Endocrinol 10:91
41. Viskum S, Rabjerg L, Jorgensen PJ et al (1999) Improvement in semen quality associated with decreasing occupational lead exposure. Am J Ind Med 35(3):257–263
42. Eibensteiner L, Del Carpio Sanz A, Frumkin H et al (2005) Lead exposure and semen quality among traffic police in Arequipa, Peru. Int J Occup Environ Health 11 (2):161–166
43. Hernandez-Ochoa I, Garcia-Vargas G, Lopez-Carrillo L et al (2005) Low lead environmental exposure alters semen quality and sperm chromatin condensation in northern Mexico. Reprod Toxicol 20(2):221–228
44. Pandey R, Singh SP (2002) Effects of molybdenum on fertility of male rats. Biometals 15 (1):65–72
45. Meeker JD, Rossano MG, Protas B et al (2008) Cadmium, lead, and other metals in relation to semen quality: human evidence for molybdenum as a male reproductive toxicant. Environ Health Perspect 116 (11):1473–1479
46. Iavicoli I, Fontana L, Bergamaschi A (2009) The effects of metals as endocrine disruptors. J Toxicol Environ Health B Crit Rev 12 (3):206–223
47. Meeker JD, Rossano MG, Protas B et al (2010) Environmental exposure to metals and male reproductive hormones: circulating testosterone is inversely associated with blood molybdenum. Fertil Steril 93(1):130–140
48. Lyubimov AV, Smith JA, Rousselle SD et al (2004) The effects of tetrathiomolybdate (TTM, NSC-714598) and copper supplementation on fertility and early embryonic development in rats. Reprod Toxicol 19 (2):223–233
49. Zhai XW, Zhang YL, Qi Q et al (2013) Effects of molybdenum on sperm quality and testis oxidative stress. Syst Biol Reprod Med 59 (5):251–255
50. Danadevi K, Rozati R, Reddy PP et al (2003) Semen quality of Indian welders occupationally exposed to nickel and chromium. Reprod Toxicol 17(4):451–456
51. Li H, Chen Q, Li S et al (2001) Effect of Cr (VI) exposure on sperm quality: human and animal studies. Ann Occup Hyg 45 (7):505–511
52. Bonde JP, Ernst E (1992) Sex hormones and semen quality in welders exposed to hexavalent chromium. Hum Exp Toxicol 11 (4):259–263
53. Bassey IE, Essien OE, Isong IKP, Udoh AE, Agbara GS (2013) Seminal plasma chromium, cadmium and lead levels in infertile men. J Med Sci 13:497–500
54. Kumar S, Sathwara NG, Gautam AK et al (2005) Semen quality of industrial workers occupationally exposed to chromium. J Occup Health 47(5):424–430
55. Nair N, Bedwal S, Prasad S et al (2005) Short-term zinc deficiency in diet induces increased oxidative stress in testes and epididymis of rats. Indian J Exp Biol 43(9):786–794
56. Aitken RJ, Roman SD (2008) Antioxidant systems and oxidative stress in the testes. Adv Exp Med Biol 636:154–171
57. Gregotti C, Di Nucci A, Costa LG et al (1992) Effects of thallium on primary cultures of testicular cells. J Toxicol Environ Health 36 (1):59–69
58. Formigli L, Scelsi R, Poggi P et al (1986) Thallium-induced testicular toxicity in the rat. Environ Res 40(2):531–539
59. Hoffman RS (2000) Thallium poisoning during pregnancy: a case report and comprehensive literature review. J Toxicol Clin Toxicol 38(7):767–775
60. Meistrich ML (1984) Stage-specific sensitivity of spermatogonia to different chemotherapeutic drugs. Biomed Pharmacother 38 (3):137–142
61. Meistrich ML, Wilson G, Brown BW et al (1992) Impact of cyclophosphamide on long-term reduction in sperm count in men treated with combination chemotherapy for

Ewing and soft tissue sarcomas. Cancer 70 (11):2703–2712

62. Meistrich ML (2009) Male gonadal toxicity. Pediatr Blood Cancer 53(2):261–266
63. Howell SJ, Shalet SM (2001) Testicular function following chemotherapy. Hum Reprod Update 7(4):363–369
64. Gerl A, Muhlbayer D, Hansmann G et al (2001) The impact of chemotherapy on Leydig cell function in long term survivors of germ cell tumors. Cancer 91(7):1297–1303
65. Petersen PM, Giwercman A, Skakkebaek NE et al (1998) Gonadal function in men with testicular cancer. Semin Oncol 25 (2):224–233
66. Howell SJ, Radford JA, Ryder WD et al (1999) Testicular function after cytotoxic chemotherapy: evidence of Leydig cell insufficiency. J Clin Oncol 17(5):1493–1498
67. Hughes CM, Lewis SE, McKelvey-Martin VJ et al (1996) A comparison of baseline and induced DNA damage in human spermatozoa from fertile and infertile men, using a modified comet assay. Mol Hum Reprod 2 (8):613–619
68. De Iuliis GN, Newey RJ, King BV et al (2009) Mobile phone radiation induces reactive oxygen species production and DNA damage in human spermatozoa in vitro. PLoS One 4(7): e6446
69. Kumar D, Salian SR, Kalthur G et al (2013) Semen abnormalities, sperm DNA damage and global hypermethylation in health workers occupationally exposed to ionizing radiation. PLoS One 8(7):e69927
70. Colpi GM, Contalbi GF, Nerva F et al (2004) Testicular function following chemo-radiotherapy. Eur J Obstet Gynecol Reprod Biol 113(Suppl 1):S2–S6
71. Knapp CA, Quinn GP, Murphy D (2011) Assessing the reproductive concerns of children and adolescents with cancer: challenges and potential solutions. J Adolesc Young Adult Oncol 1(1):31–35
72. Stern C, Conyers R, Orme L et al (2013) Reproductive concerns of children and adolescents with cancer: challenges and potential solutions. Clin Oncol Adolesc Young Adult 3:63–78
73. Lee HR, Jeung EB, Cho MH et al (2013) Molecular mechanism(s) of endocrine-disrupting chemicals and their potent oestrogenicity in diverse cells and tissues that express oestrogen receptors. J Cell Mol Med 17 (1):1–11
74. Gray LE Jr (1998) Xenoendocrine disrupters: laboratory studies on male reproductive effects. Toxicol Lett 102–103:331–335
75. Luccio-Camelo DC, Prins GS (2011) Disruption of androgen receptor signaling in males by environmental chemicals. J Steroid Biochem Mol Biol 127(1–2):74–82
76. Defarge N, Mesnage R, Gress S et al (2012) Letter to the editor: developmental and reproductive outcomes of roundup and glyphosate in humans and animals. J Toxicol Environ Health B Crit Rev 15(7):433–437, author reply 438–440
77. Clair E, Mesnage R, Travert C et al (2012) A glyphosate-based herbicide induces necrosis and apoptosis in mature rat testicular cells in vitro, and testosterone decrease at lower levels. Toxicol In Vitro 26(2):269–279
78. de Liz Oliveira Cavalli VL, Cattani D, Heinz Rieg CE et al (2013) Roundup disrupts male reproductive functions by triggering calcium-mediated cell death in rat testis and Sertoli cells. Free Radic Biol Med 65C:335–346
79. Williams AL, Watson RE, DeSesso JM (2012) Developmental and reproductive outcomes in humans and animals after glyphosate exposure: a critical analysis. J Toxicol Environ Health B Crit Rev 15(1):39–96
80. Vonier PM, Crain DA, McLachlan JA et al (1996) Interaction of environmental chemicals with the estrogen and progesterone receptors from the oviduct of the American alligator. Environ Health Perspect 104 (12):1318–1322
81. Grizard G, Ouchchane L, Roddier H et al (2007) In vitro alachlor effects on reactive oxygen species generation, motility patterns and apoptosis markers in human spermatozoa. Reprod Toxicol 23(1):55–62
82. Swan SH, Liu F, Overstreet JW et al (2007) Semen quality of fertile US males in relation to their mothers' beef consumption during pregnancy. Hum Reprod 22(6):1497–1502
83. Lenzi A, Picardo M, Gandini L et al (1996) Lipids of the sperm plasma membrane: from polyunsaturated fatty acids considered as markers of sperm function to possible scavenger therapy. Hum Reprod Update 2 (3):246–256
84. Bongalhardo DC, Leeson S, Buhr MM (2009) Dietary lipids differentially affect membranes from different areas of rooster sperm. Poult Sci 88(5):1060–1069
85. Afeiche M, Williams PL, Mendiola J et al (2013) Dairy food intake in relation to semen quality and reproductive hormone levels among physically active young men. Hum Reprod 28(8):2265–2275
86. Safarinejad MR, Safarinejad S (2012) The roles of omega-3 and omega-6 fatty acids in

idiopathic male infertility. Asian J Androl 14 (4):514–515

87. Safarinejad MR, Hosseini SY, Dadkhah F et al (2010) Relationship of omega-3 and omega-6 fatty acids with semen characteristics, and anti-oxidant status of seminal plasma: a comparison between fertile and infertile men. Clin Nutr 29(1):100–105
88. Castellano CA, Audet I, Bailey JL et al (2010) Effect of dietary n-3 fatty acids (fish oils) on boar reproduction and semen quality. J Anim Sci 88(7):2346–2355
89. Yan L, Bai XL, Fang ZF et al (2013) Effect of different dietary omega-3/omega-6 fatty acid ratios on reproduction in male rats. Lipids Health Dis 12:33
90. Wathes DC, Abayasekara DR, Aitken RJ (2007) Polyunsaturated fatty acids in male and female reproduction. Biol Reprod 77 (2):190–201
91. Zhu Q, Van Thiel DH, Gavaler JS (1997) Effects of ethanol on rat Sertoli cell function: studies in vitro and in vivo. Alcohol Clin Exp Res 21(8):1409–1417
92. Muthusami KR, Chinnaswamy P (2005) Effect of chronic alcoholism on male fertility hormones and semen quality. Fertil Steril 84 (4):919–924
93. Gordon GG, Altman K, Southren AL et al (1976) Effect of alcohol (ethanol) administration on sex-hormone metabolism in normal men. N Engl J Med 295(15):793–797
94. Emanuele MA, Emanuele NV (1998) Alcohol's effects on male reproduction. Alcohol Health Res World 22(3):195–201
95. Gaur DS, Talekar MS, Pathak VP (2010) Alcohol intake and cigarette smoking: impact of two major lifestyle factors on male fertility. Indian J Pathol Microbiol 53(1):35–40
96. Ouko LA, Shantikumar K, Knezovich J et al (2009) Effect of alcohol consumption on CpG methylation in the differentially methylated regions of H19 and IG-DMR in male gametes: implications for fetal alcohol spectrum disorders. Alcohol Clin Exp Res 33 (9):1615–1627
97. Bielawski DM, Zaher FM, Svinarich DM et al (2002) Paternal alcohol exposure affects sperm cytosine methyltransferase messenger RNA levels. Alcohol Clin Exp Res 26 (3):347–351
98. Practice Committee of the American Society for Reproductive M (2012) Smoking and infertility: a committee opinion. Fertil Steril 98(6):1400–1406
99. Stephens WE, Calder A, Newton J (2005) Source and health implications of high toxic metal concentrations in illicit tobacco products. Environ Sci Technol 39(2):479–488
100. Al-Bader A, Omu AE, Dashti H (1999) Chronic cadmium toxicity to sperm of heavy cigarette smokers: immunomodulation by zinc. Arch Androl 43(2):135–140
101. Meri ZB, Irshid IB, Migdadi M et al (2013) Does cigarette smoking affect seminal fluid parameters? A comparative study. Oman Med J 28(1):12–15
102. Yamamoto Y, Isoyama E, Sofikitis N et al (1998) Effects of smoking on testicular function and fertilizing potential in rats. Urol Res 26(1):45–48
103. Sofikitis N, Miyagawa I, Dimitriadis D et al (1995) Effects of smoking on testicular function, semen quality and sperm fertilizing capacity. J Urol 154(3):1030–1034
104. Fraga CG, Motchnik PA, Wyrobek AJ et al (1996) Smoking and low antioxidant levels increase oxidative damage to sperm DNA. Mutat Res 351(2):199–203
105. Zenzes MT, Puy LA, Bielecki R et al (1999) Detection of benzo[a]pyrene diol epoxide-DNA adducts in embryos from smoking couples: evidence for transmission by spermatozoa. Mol Hum Reprod 5(2):125–131
106. Yu B, Qi Y, Liu D et al (2014) Cigarette smoking is associated with abnormal histone-to-protamine transition in human sperm. Fertil Steril 101(1):51–57.e51
107. Marczylo EL, Amoako AA, Konje JC et al (2012) Smoking induces differential miRNA expression in human spermatozoa: a potential transgenerational epigenetic concern? Epigenetics 7(5):432–439
108. el Mohamed SA, Song WH, Oh SA et al (2010) The transgenerational impact of benzo(a)pyrene on murine male fertility. Hum Reprod 25(10):2427–2433
109. Prakash N, Vijay KM, Sunilchandra U et al (2010) Evaluation of testicular toxicity following short-term exposure to cypermethrin in albino mice. Toxicol Int 17(1):18–21
110. Ige SF, Olaleye SB, Akhigbe RE et al (2012) Testicular toxicity and sperm quality following cadmium exposure in rats: ameliorative potentials of Allium cepa. J Hum Reprod Sci 5(1):37–42
111. Creasy DM (2001) Pathogenesis of male reproductive toxicity. Toxicol Pathol 29 (1):64–76
112. Misell LM, Holochwost D, Boban D et al (2006) A stable isotope-mass spectrometric method for measuring human spermatogenesis kinetics in vivo. J Urol 175(1):242–246, discussion 246

113. Clifton DK, Bremner WJ (1983) The effect of testicular x-irradiation on spermatogenesis in man - a comparison with the mouse. J Androl 4(6):387–392
114. Mangochi P (2010) Endocrine distrupting chemicals and human health: the plausibility of research results on DDT and reproductive health. Malawi Med J 22(2):42–45
115. Dere E, Anderson LM, Coulson M et al (2013) SOT symposium highlight: translatable indicators of testicular toxicity: inhibin B, microRNAs, and sperm signatures. Toxicol Sci 136(2):265–273
116. Elkin ND, Piner JA, Sharpe RM (2010) Toxicant-induced leakage of germ cell-specific proteins from seminiferous tubules in the rat: relationship to blood-testis barrier integrity and prospects for biomonitoring. Toxicol Sci 117(2):439–448
117. Anton E, Krawetz SA (2012) Spermatozoa as biomarkers for the assessment of human male infertility and genotoxicity. Syst Biol Reprod Med 58(1):41–50
118. Amaral S, Amaral A, Ramalho-Santos J (2013) Aging and male reproductive function: a mitochondrial perspective. Front Biosci (Schol Ed) 5:181–197
119. Swan SH, Brazil C, Drobnis EZ et al (2003) Geographic differences in semen quality of fertile U.S. males. Environ Health Perspect 111(4):414–420
120. Vine MF, Margolin BH, Morrison HI et al (1994) Cigarette smoking and sperm density: a meta-analysis. Fertil Steril 61(1):35–43
121. Check JH, Epstein R, Long R (1991) Effect of time interval between ejaculations on semen parameters. Arch Androl 27(2):93–95
122. Laterza OF, Lim L, Garrett-Engele PW et al (2009) Plasma microRNAs as sensitive and specific biomarkers of tissue injury. Clin Chem 55(11):1977–1983
123. Sakkas D, Alvarez JG (2010) Sperm DNA fragmentation: mechanisms of origin, impact on reproductive outcome, and analysis. Fertil Steril 93(4):1027–1036
124. Meyer JN, Leung MCK, Rooney JP et al (2013) Mitochondria as a target of environmental toxicants. Toxicol Sci 134(1):1–17
125. De Mas P, Daudin M, Vincent MC et al (2001) Increased aneuploidy in spermatozoa from testicular tumour patients after chemotherapy with cisplatin, etoposide and bleomycin. Hum Reprod 16(6):1204–1208
126. Sendler E, Johnson GD, Mao S et al (2013) Stability, delivery and functions of human sperm RNAs at fertilization. Nucleic Acids Res 41(7):4104–4117
127. Jodar M, Kalko S, Castillo J et al (2012) Differential RNAs in the sperm cells of asthenozoospermic patients. Hum Reprod 27 (5):1431–1438
128. Platts AE, Dix DJ, Chemes HE et al (2007) Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. Hum Mol Genet 16(7):763–773
129. Pacheco SE, Anderson LM, Sandrof MA et al (2012) Sperm mRNA transcripts are indicators of sub-chronic low dose testicular injury in the Fischer 344 rat. PLoS One 7(8): e44280
130. Amaral A, Castillo J, Ramalho-Santos J et al (2014) The combined human sperm proteome: cellular pathways and implications for basic and clinical science. Hum Reprod Update 20(1):40–62
131. Auger J, Eustache F, Maceiras P et al (2010) Modified expression of several sperm proteins after chronic exposure to the antiandrogenic compound vinclozolin. Toxicol Sci 117 (2):475–484
132. Maselli J, Hales BF, Chan P et al (2012) Exposure to bleomycin, etoposide, and cis-platinum alters rat sperm chromatin integrity and sperm head protein profile. Biol Reprod 86(5):166, 161–110
133. Li Y, Okumura K, Nomura S et al (2011) Oxidatively damaged proteins in the early stage of testicular toxicities in male rats by orally administered with a synthetic oestrogen, diethylstilbestrol. Reprod Toxicol 31 (1):26–34
134. Martinez-Heredia J, de Mateo S, Vidal-Taboada JM et al (2008) Identification of proteomic differences in asthenozoospermic sperm samples. Hum Reprod 23(4):783–791
135. Creasy DM, Ford GR, Gray TJ (1990) The morphogenesis of cyclohexylamine-induced testicular atrophy in the rat: in vivo and in vitro studies. Exp Mol Pathol 52 (2):155–169
136. Creasy DM, Foster PM (1984) The morphological development of glycol ether-induced testicular atrophy in the rat. Exp Mol Pathol 40(2):169–176
137. Hess RA, Linder RE, Strader LF et al (1988) Acute effects and long-term sequelae of 1,3-dinitrobenzene on male reproduction in the rat II. Quantitative and qualitative histopathology of the testis. J Androl 9(5):327–342
138. Blackburn DM, Gray AJ, Lloyd SC et al (1988) A comparison of the effects of the three isomers of dinitrobenzene on the testis in the rat. Toxicol Appl Pharmacol 92 (1):54–64

139. Kerr JB, Millar M, Maddocks S et al (1993) Stage-dependent changes in spermatogenesis and Sertoli cells in relation to the onset of spermatogenic failure following withdrawal of testosterone. Anat Rec 235(4):547–559
140. Chapin RE, Morgan KT, Bus JS (1983) The morphogenesis of testicular degeneration induced in rats by orally administered 2,5-hexanedione. Exp Mol Pathol 38 (2):149–169
141. Lambrot R, Muczynski V, Lecureuil C et al (2009) Phthalates impair germ cell development in the human fetal testis in vitro without change in testosterone production. Environ Health Perspect 117(1):32–37
142. Ernst E (1990) Testicular toxicity following short-term exposure to tri- and hexavalent chromium: an experimental study in the rat. Toxicol Lett 51(3):269–275
143. Garg S, Doncel G, Chabra S et al (1994) Synergistic spermicidal activity of neem seed extract, reetha saponins and quinine hydrochloride. Contraception 50(2):185–190
144. Creasy DM (2011) Commentary on incidence and nature of testicular toxicity findings in pharmaceutical development survey: a pathologist's perspective. Birth Defects Res B Dev Reprod Toxicol 92(6):508–510
145. Gray TJ (1986) Testicular toxicity in vitro: Sertoli-germ cell co-cultures as a model system. Food Chem Toxicol 24(6–7):601–605
146. Emanuele MA, Emanuele N (2001) Alcohol and the male reproductive system. Alcohol Res Health 25(4):282–287
147. Sarkar R, Mohanakumar KP, Chowdhury M (2000) Effects of an organophosphate pesticide, quinalphos, on the hypothalamo-pituitary-gonadal axis in adult male rats. J Reprod Fertil 118(1):29–38
148. Sasaki JC, Chapin RE, Hall DG et al (2011) Incidence and nature of testicular toxicity findings in pharmaceutical development. Birth Defects Res B Dev Reprod Toxicol 92 (6):511–525
149. Dineshkumar D, Selvaraju S, Parthipan S et al (2013) Effect of detoxified karanja (Pongamia spp.) cake on testicular architecture and semen production in ram lambs. Animal 7 (10):1697–1703
150. Reader SCJ, Shingles C, Stonard MD (1991) Acute testicular toxicity of 1,3-dinitrobenzene and ethylene-glycol monomethyl ether in the rat - evaluation of biochemical effect markers and hormonal responses. Fundam Appl Toxicol 16(1):61–70
151. Matsuyama T, Niino N, Kiyosawa N et al (2011) Toxicogenomic investigation on rat testicular toxicity elicited by 1,3-dinitrobenzene. Toxicology 290(2–3): 169–177
152. Shen RS, Lee IP (1984) Selected testicular enzymes as biochemical markers for procarbazine-induced testicular toxicity. Arch Toxicol 55(4):233–238
153. Yan LY, Yue DB, Luo HL et al (2010) Effect of Vitamin E supplementation on the enzymatic activity of selected markers in Aohan fine-wool sheep testis. Anim Reprod Sci 122 (3–4):264–269
154. Turner KJ, McKinnell C, McLaren TT et al (1996) Detection of germ cell-derived proteins in testicular interstitial fluid: potential for monitoring spermatogenesis in vivo. J Androl 17(2):127–136

Methods in Pharmacology and Toxicology (2017) 205–234
DOI 10.1007/7653_2017_69
© Springer Science+Business Media New York 2017
Published online: 16 May 2017

Effects of Chemicals on Mammary Gland Development

Adam J. Filgo and Ali S. Faqi

Abstract

The mammary gland is exceptionally a complex tissue. It is a sexually dimorphic organ in function, size, response to hormone signaling, and cellular structure. Unlike most organs, the mammary gland has several critical periods of growth and development after birth, and is only fully developed after a full-term pregnancy. Mammary gland development is dependent on complex endocrine signaling as well as paracrine and autocrine signaling between the stroma and parenchyma cells. Even outside of the critical windows of growth and development, the mammary gland is constantly changing with normal hormone fluctuations, most notably during the estrous/menstrual cycle. It is particularly sensitive to endocrine disrupting chemicals (EDCs). An EDC can affect both females and males, resulting in abnormal mammary gland development in adolescents. Later in life, EDCs can influence cancer outcomes. In adult females, alterations in mammary gland development can result in lactational impairment. This chapter describes the stages of development, the key hormone actions, and common EDCs and their effects on the mammary gland.

Keywords Endocrine disrupting chemicals, Gestational development, Hormones, Involution, Lactation, Mammary gland, Pregnancy, Puberty

1 Introduction

The mammary glands of both humans and rodents undergo four phases of critical growth and development: gestational, postnatal, puberty, and pregnancy followed by involution (Fig. 1). Little is known about the mechanism of fetal breast formation in humans, but in rodents mammary gland development involves communication between the epidermis and mesenchyme by various signaling pathways [1–3]. Male and female rodents have morphologically different mammary glands in utero, while human mammary glands are not sexually dimorphic until puberty. Male mice have a 2-day surge of fetal androgens at embryonic day (E) 13.5–15.5 that causes apoptosis of the mammary epithelium and regression may eliminate the mammary epithelium. In male rats, due to the androgen surge that occurs around E17 the epithelial bud becomes separated from the epidermis and no nipples are formed, but the rudimentary ductal tree is intact. In humans there is evidence that maternal hormones play a part in growth and development of the mammary glands of fetuses and newborns. While not found in rodents, lobular structures have

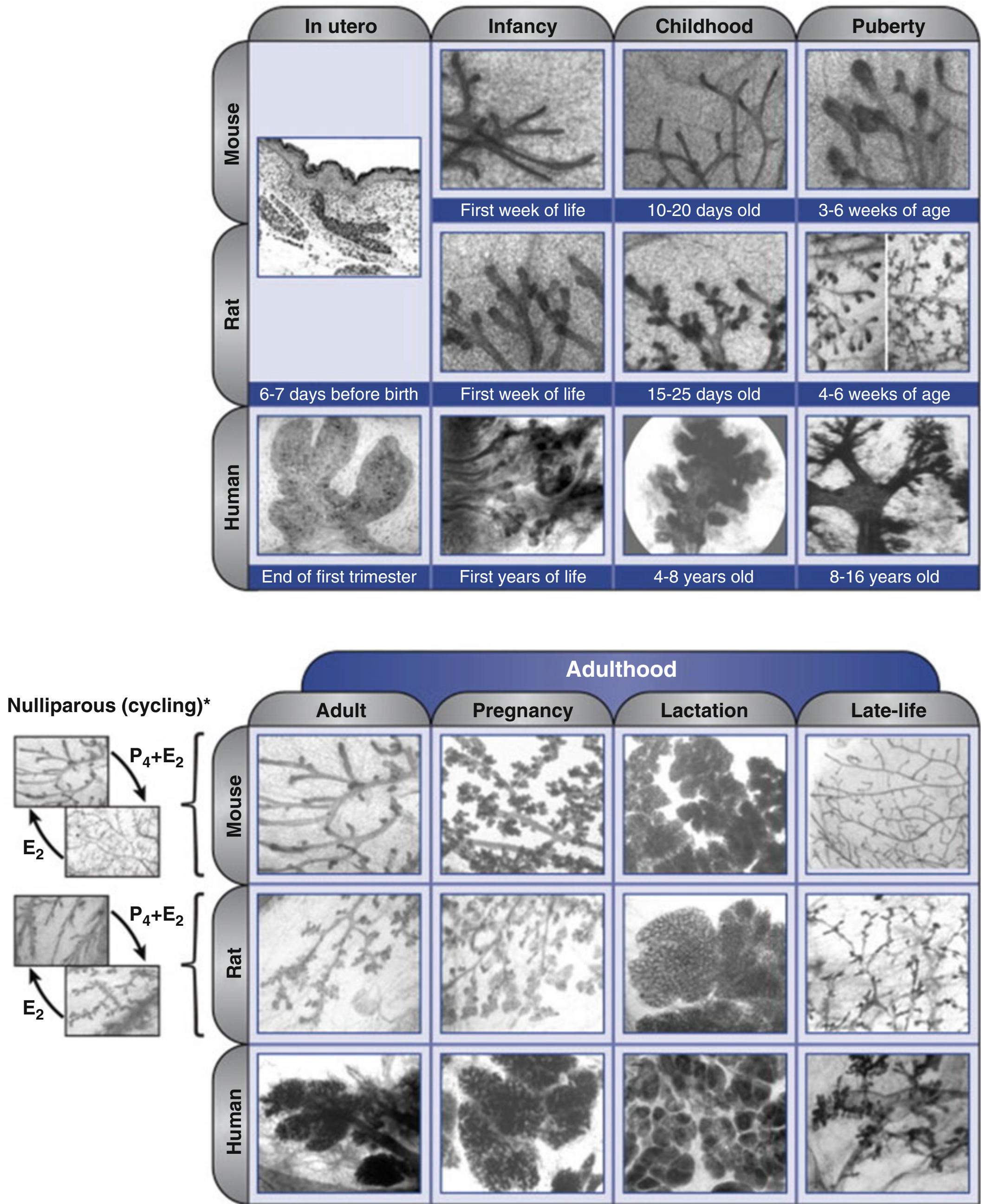

Fig. 1 Structure of the mammary gland during different life stages of the mouse, rat, and human. In this comparison of the rodent mammary gland and human breast over the life course, mouse and rat tissues are magnified 1.8 times and human tissue is magnified 2.5 times (rat tissues at 4–6 weeks are less magnified than mouse and human tissues at the same age). Source: Interagency Breast Cancer and Environmental Research Coordinating Committee, Prioritizing Prevention, 2013. http:/www.niehs.nih.gov.prox.lib.ncsu.edu/about/boards/ibcercc/

been seen in newborn infants, and are transient structure that dissipates weeks after birth.

Before the onset of puberty, isometric mammary gland growth continues in both rodents and humans, independent of hormones. Allometric, or exponential, mammary growth during puberty is dependent on several hormones and paracrine signaling [4]. In rodents, allometric growth is initiated by ovarian steroids. Estrogen and the growth factors triggered by ER activation signal ductal elongation. Ductal branching and alveolar budding are signaled by progesterone [5, 6]. In humans, regulation of allometric growth is less clear. Acini form on the terminal ducts and embed in intralobular stoma to form terminal duct lobular units (TDLUs), which are the functional units of the breast [7]. Both endocrine and paracrine signaling may also stimulate mammary gland growth in humans, but specific effects are uncertain [6]. Male mammary glands in humans normally do not develop further during puberty [7]. During pregnancy, in both humans and rodents there is an increase in the number of lobules and a loss of fat in adipose cells of the mammary gland during pregnancy and lactation [7]. Lactational involution of the mammary gland occurs after weaning of the offspring. After weaning, the TDLUs decrease in number and size, but the ducts are not involved. During the type of involution that occurs with aging (and not lactational), both the TLDUs and the ducts are reduced in number and are replaced with collagen and fat [7]. The mammary gland is one of the few truly dynamic organs and undergoes many periods of proliferation and remodeling. For this reason, carcinogens and endocrine disruption chemicals (EDCs) are thought to target the mammary gland (Fig. 2). The timing and length of these growth periods may vary across species and may affect carcinogen susceptibility. Further investigation of the susceptibility of the mammary gland during periods of growth and development is needed to understand lifelong breast cancer risk.

Breast cancer is the most commonly diagnosed cancer in women, and after lung cancer, is the second most common fatal cancer of women [9]. In 2016, there was an estimated 246,660 new cases of invasive breast cancer diagnosed in women in the USA, which equates to about one in every eight women [10]. The peak age range for diagnosis of breast cancer in women is 60–69 [11]. Between 2002 and 2003, there was a decrease in the breast cancer incidence rate of almost 7% due to reductions in use of hormonal replacement therapy in postmenopausal women [10]. Between 2004 and 2013, breast cancer incidence rates have remained stable in White, American Indian/Alaskan Native, and Hispanic women and slightly increased in Black and Asian/Pacific Islander women (by 0.6% and 0.9% per year, respectively) [9]. Deaths from breast cancer have decreased by 1.9% in White women and 1.4% in Black women from 2003 to 2012 [10]. Breast cancer in men is uncommon and accounts for about 1% of total breast cancers [11]. In 2016, an estimated 2,600 men will be diagnosed with invasive breast cancer in the USA [10]. Unlike breast cancer in women,

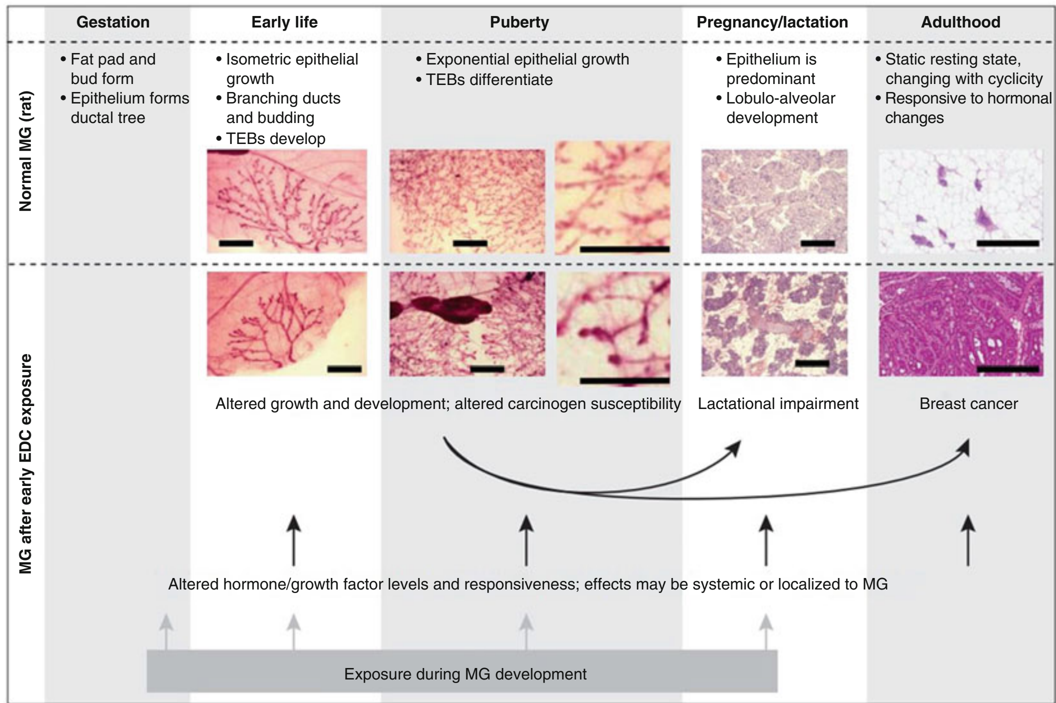

Fig. 2 Stages of normal rat mammary gland (MG) development and effects of environment on subsequent events. Effects of early life EDC exposures can lead to altered developmental programming in the breast and have been reported neonatally, at puberty, and well into adulthood, when effects on lactation or mammary tumorigenesis become evident. *Arrows* indicate plausible (*black*) or more certain (*gray*) mechanistic pathways. Photomicrographs for early life and puberty were all taken at 16× magnification; photomicrographs for pregnancy/lactation and adulthood were taken at 10× magnification. Bars = 2 mm. Source: Rudel et al. [8]

male breast cancer incidence rates may be slowly rising [11]. Most diagnoses of breast cancer, in both women and men, are without associated breast cancer susceptibility gene mutations [11].

Known risk factors for women include family history, genetic variants (BRCA1, BRCA2, TP53, PTEN, etc.), breast density, history of benign breast disease, reproductive factors (increased lifetime exposures to estrogen and progesterone and not having children), physical activity, alcohol consumption, radiation exposure, and adult body mass [12]. Potential risk factors for men include family history, estrogen excess (medication, Klinefelter syndrome), genetic variants (similar to those in women), occupational exposures to radiation, carcinogens, and potentially volatile organic compounds [12]. Environmental and lifestyle factors contribute to 70–95% of breast cancer risk [13]. Even familial breast cancer risk is speculated to be the interaction between lifestyle factors and common gene variations (that only add a small increase in breast cancer risk) [10].

Exogenous chemical exposures (aryl aromatic amines, N-nitrosamines, tobacco smoke, etc.) have had significant correlations with

risk for breast cancer development in epidemiological studies around the globe [12]. However, it is difficult to prove causation because the disease is often diagnosed decades after environmental exposures and humans are not exposed to one chemical. Even when causation may be attributed to a particular environmental exposure, such as in rodent model studies in a controlled environment, it is often complicated to determine a mechanism by which it affects the mammary tissue due to the complexity of the gland and its intricate signaling pathways. There are windows of susceptibility in mammary gland development, not common to other organs, when exogenous chemical exposures could influence breast development and increase breast cancer risk [14].

2 Key Stages of Mammary Gland Development

2.1 Gestation

The hallmarks of human embryonic mammary gland development are typically reported in weeks of gestation and have been extensively reviewed in the literature and are as follows [6, 15–22]. Female and male breast development is indistinguishable throughout gestation. There are variations on the exact time based on how the authors define the beginning of gestation (estimated time of conception, last missed menstrual period, or length of the embryo). The first key stage of mammary gland development begins as early as week 4 of gestation with the development of the "milk line," lines of thickening in the ectoderm that reaches from the groin to the axilla on the ventrolateral area of the embryo. The ectoderm is surrounded by a basement membrane, separating it from the mesoderm. By week 8, the ectoderm grows into the unevenly vascularized and specialized mesoderm and continues to thicken to form a four to six cell wide bilateral ridge or "mammary crest" in the thoracic region, and the rest of the milk line starts to involute. The ridges continue to regresses and form one pair of mammary placodes, which have a button-like appearance on the surface of the ectoderm. Supernumerary nipples, however, may develop at any location along the milk line. By week 12, the ectodermal cells of a mammary placode continue to aggregate and form the "breast bud." The flask shaped breast bud is composed of two distinct cell populations, a central and a peripheral one. The mesoderm surrounding the breast bud differentiates to form fibroblasts, smooth muscle cells, capillary endothelial cells, and adipocytes. The ectoderm grows inward and then the surrounding mesoderm proliferates to create the nipples and areolae between weeks 12 and 16. The nipples and areolae are composted of smooth muscle fibers oriented both circularly and longitudinally. The breast bud does not change in appearance until week 20. Then the breast bud indents to form secondary buds and projections sprout from the secondary buds to form 15–25 secondary projections. The projections have a slender stalk and club-like end that will grow through the mesenchymal into the

stroma. The secondary projections then begin to canalize via apoptosis to form primary lactiferous ducts by week 32. The cells lining the lactiferous ducts are arranged in two layers, the luminal cells and myoepithelial cells. A mammary pit forms and the lactiferous ducts drain into ampullae that open into the mammary pit. By birth, the areola is formed and a varying amount of lobuloalveolar development of the mammary gland may occur.

2.2 Postnatal and Prepubertal

After birth, the mesoderm underneath the nipples proliferates, causing the nipples to evert [22]. Inverted nipples are a common phenomenon (up to 20% of women) in which the nipple fails to evert [23]. This is a benign clinical finding but may affect future breastfeeding in females. Areolae pigmentation increases and the Montgomery tubercles (coalesced sebaceous glands and lactiferous units that protect and lubricate the nipples and areolae) can be seen on the surface of the areolae [24]. In the first 2 years of life there is a variable amount of lobular development found in both females and males. Both rudimentary mammary ducts with club-like terminal ends and type 1 lobules similar to those found in adults can be found in the infant breast. Involution of lobular development occurs by 2 years and then the mammary gland remains relatively quiescent until puberty [22, 25]. Between involution and puberty the breast grows isometrically and there are no distinguishable differences between male and female mammary glands [21].

2.3 Puberty

Sexually dimorphic breast development occurs at the start of puberty, due to the rise in sex hormones levels. There is a significant variance in the onset of female breast development (thelarche) with a normal range of 8.5–13 years [22]. Once thelarche has begun, the female breast begins to grow exponentially as described macroscopically by the Tanner stages. Stage 1 is the prepubertal breast. Stage 2 is the breast bud stage with elevation of breast and papilla and the diameter of the areola enlarges. Stage 3 is further enlargement of the breast and areola, without separation of the areola contour from the surrounding breast. Stage 4 is when areola and papilla form a secondary mound above the level of the breast. Stage 5 is the mature stage and the areola returns to the general contour of the breast [22, 26]. The average time spent between the onset of thelarche (Tanner stage 2) and completion (Tanner stage 5) is 4–4½ years, with a considerable amount of normal variability [22]. In males, no further development of the breast occurs due to increased testosterone concentrations [22].

The mammary gland epithelium and surround stroma undergo extensive growth and differentiation. As the stroma expands with increasing connective and fatty tissue, the mammary epithelium follows with ductal elongation and branching [16]. The site of the elongation and branching occurs at the terminal end buds (TEB), the proliferation that occurs in TEBs are due to the population of resident mammary

stem and progenitor cells [16, 21, 27–29]. The primary ducts give rise to segmental ducts and to smaller subsegmental ducts. The subsegmental ducts lead to terminal ducts and acini [22]. The functional unit of the breast is made up of acini from one terminal duct embedded in the stroma and is referred to as a terminal duct lobular unit (TDLU). Mammary gland epithelium proliferation increases and lobule type 1 form. [6] Each lobule type 1 is composed of about 11 alveolar buds and is the predominate lobule in nulliparous women. Full differentiation of the mammary gland does not occur without pregnancy. [30] Lobule types 2 and 3 (containing on average of 47 and 81 alveolar buds, respectively) are present in nulliparous women and gradually increase with each menstrual cycle [16].

2.4 Pregnancy and Lactation

During pregnancy the breast increases in water, electrolyte, and fat content. This increase in overall breast volume is accompanied by a twofold increase in the vascular supply. Size and pigmentation of the nipples increase and the Montgomery tubercles greatly enlarge [31]. The mammary gland epithelium undergoes another round of growth and differentiation, similar to that of puberty. Unlike during puberty, the increase in the number of lobules is at the expense of the surrounding stroma and fat. Proliferation of distal alveolar buds transforms type 1 lobules through type 3 lobules and continues in the formation of acini found in type 4 lobules. The secretory alveoli not only increase in number, but also in size due to cytoplasm enlargement. The mammary gland epithelial architecture is essentially settled by the first half of pregnancy. The continued growth of the mammary gland in the second half of pregnancy is due to the increased height of the epithelial cells and expansion of the lumen of the acini becoming distended with secretory material [32]. During lactation the TDLU are enlarged and the dilated lumens are filled with milk. As long as the milk is regularly removed from the mammary gland milk secretion will continue.

2.5 Involution

There are two periods of major mammary gland involution: post-lactational and post-menopausal. Post-lactational involution occurs after cessation of milk removal from the breast (weaning). After weaning, milk production and synthesis stop quickly and the first phase of involution involved apoptosis and regression of the epithelial and stromal tissues of the breast. Phase I is reversible and lactation can be restarted by removal of the milk from the breast. Phase II is irreversible and is characterized by luminal cell loss, stromal remodeling, and a regeneration of the connective tissue [33]. Once post-lactational involution has completed the mammary gland has regressed back to its resting state. However, the parous gland has considerably more glandular tissue, primarily comprised of lobule type 2 and 3 units, than the nulliparous gland. More PRA is expressed in parous women compared to nulliparous women, with no significant difference in ERα or PRB [34].

Post-menopausal involution begins after the fourth decade of life. Lobule type 2 and 3 TDLUs gradually regress to lobule type 1. After menopause this regression is substantially increased and the percentage of lobule type 1 present in the breast is equal between post-menopausal parous and nulliparous women. Although both parous and nulliparous women have predominantly type 1 lobules, parous women have increased intralobular hyalinization and decreased cellular proliferation compared to nulliparous women [35].

3 Hormone Actions in Mammary Gland Development

3.1 Gestation

During gestation mammary gland development is largely sex hormone independent. The known signaling events of embryonic mammary gland development from the formation of the milk line to the rudimentary ductal tree present at birth have been described primarily in rodents. Some of the signaling molecules and pathways mediating the developmental progression of the mammary gland during embryogenesis include Wnt/β-catenin, Lef/Tcf, MSX1/MSX2, PTHrP, keratinocyte growth factor, neuregulin, activins, and inhibins [2, 18, 32, 36, 37]. These signaling molecules are found either in the epithelium, the mesenchyme or both and thus may mediate endocrine, paracrine and/or autocrine influences on mammary epithelial growth. The canonical Wnt signaling pathway and the fibroblast growth factor (fgf) signaling pathway activate TBX3 expression and result in the formation of the milk line. TBX3 expression induces the expression of specific Wnts and fgfs that act in an autocrine fashion and lead to the formation of the mammary placodes [2]. Wnt and the subsequent stabilization of β-catenin enter the nucleus together with Tcf/Lef to directly promote transcription, possibly by modulating cell adhesion and promoting cell migration in the mammary placode [36, 38, 39]. The mammary bud then forms due to a number of signaling pathways within either the epithelial or mesenchymal cells including PTHrP, Wnt, and IGF1 and transcription factors Msx1/Msx2 [2]. Msx1/Msx2 are necessary, but redundant, in the formation of the mammary buds. PTHrP is produced by the breast bud and will trigger adjacent mesenchymal cells to acquire a specialized mammary fate and to develop the nipple and areola [2]. PTHrP is androgen-regulated and in rodents, an androgen surge during embryogenesis will inhibit the formation of nipples/areolae in male rodents. The mammary gland continues to develop to form a rudimentary ductal tree. PTHrP, BMP, and Wnt trigger and/or sustain ductal outgrowth from the mammary buds [36]. The hormone receptors involved in mammary gland development and differentiation during puberty including estrogen receptors (ER) α and β, progesterone receptors, prolactin receptors, and growth hormone (GH) receptors are not necessary during gestational development of the mammary gland [2, 36].

3.2 Postnatal and Prepubertal

At birth, maternally derived prolactin level can be high enough in the newborn that the mammary glands to produce colostrum. As the maternally derived prolactin in the newborn diminishes, the mammary gland regresses and production of colostrum ceases. [6] The mammary gland will remain quiescent until the onset of puberty. Transforming growth factors (α and β), tenascin-C and collagen type 4 found in the epithelium, and the surrounding stroma might play important role in differentiation and growth in early human breast development. TGFβ in the infant breasts might play an important role in inhibiting the epithelial proliferation at specific sites [40]. The epithelial growth factor (EGF) receptor (EGFR) has been identified as a mediator of epithelial–stromal interactions and EGF or its relatives play a role during isometric growth of the rudimentary mammary. Other pathways believed to be involved in isometric growth of the mammary gland include ERα, ERBB2, insulin like growth factor-1 (IGF) prolactin, and growth hormone.

3.3 Puberty

At the onset of thelarche changes in the pulses of gonadotropin-releasing hormone from the hypothalamus leads to increasing levels of circulating sex hormones, estradiol, androstenedione, androsterone, and gonadotropins [15]. Both males and females have increasing levels of estrogens and androgens during puberty. Mammary development is typically completely suppressed in males by the elevated ratio of androgens to estrogens. For mammary gland proliferation the hormones estrogen, progesterone, and prolactin or growth hormone are essential for normal development. [6, 41] It is possible that estrogen also cooperates with LH and FSH to effect normal breast development Ductal elongation in the female mammary gland is primarily influenced by estrogens, GH, and IGF-1. In the rat, estradiol cannot stimulate TEBs formation in the absence of GH or IGF-I, while GH or IGF-I can only be partially stimulatory in the absence of estradiol [42]. Epithelial proliferation in the TEBs is almost responsible for ductal outgrowth and relies on the presence of ERα. Stromal interactions with estrogens do not play an essential role in epithelial proliferation and ductal outgrowth. ERα expression in the mammary epithelium is heterogeneous and estrogens act on the ERα positive epithelial cells causing them to release paracrine signals that allow ERα negative epithelial cells to proliferate. In response to increased levels of estradiol the ERα positive epithelial cells release amphiregulin, an EGFR ligand, resulting in the formation of TEBs, proliferation in surrounding ERα negative epithelial cells and ductal outgrowth [43, 44]. Mice lacking members of the TGFβ family, activins and inhibins fail to undergo ductal elongation during puberty [18, 37].

Progesterone is able to induce proliferation both directly on progesterone receptor (PR) positive cells via cyclin D1 signaling pathway and indirectly on PR negative cells through the paracrine mediator amphiregulin [1, 45]. Like estrogen, progesterone is able to activate TEB formation during puberty. Progesterone is essential for side branching

and alveologenesis of the mammary gland. Receptor activator of NF-κB ligand (RANKL) is a downstream effector of progesterone in the mammary gland and is responsible for progesterone-induced side branching and alveolar budding [46]. Wnt4 is a downstream target of progesterone and activates proliferation target genes [45, 47]. Progesterone targets such as RANKL, Wnt4, and amphiregulin that have been characterized as paracrine mediators in the rodent mammary gland were present in human mammary epithelial subpopulations but not upregulated in vitro [45].

Allometric mammary gland development is largely dependent on growth hormone [48]. Paracrine signaling of IGF-1 activated the GHR to produce GH. GHR knockout mice have delayed ductal outgrowth [41]. Information concerning the contribution of growth hormone, IGF, and prolactin to allometric mammary development in girls is lacking, although it is clear that serum levels of these hormones increase in girls during puberty and may therefore contribute to aspects of TDLU development. LEF-1, hepatocyte growth factor, keratin growth factor, and neuregulin are produced by the mesenchyme. Studies have shown that transforming growth factor β (TGFβ) plays a role in impeding proliferation of ER/PR positive cells [49]. Vitamin D receptor serves as a negative growth regulator by suppression of branching morphogenesis during puberty [32].

During the menstruation cycle, epithelial proliferation increases. Proliferation of epithelial cells in the breast correlates with serum progesterone levels during the menstrual cycle and is maximal during the luteal phase, peaking at days 23–25. This increase in proliferation also coincides with a peak in estradiol [50]. Despite the clear effect of progesterone, this increased proliferation appears to be due, at least in part, to estradiol. Antiestrogenic and antiproliferative effects of progesterone on breast mammary epithelial cells have also been documented, indicating that a delicate balance of estradiol and progesterone directs normal breast development. Based on studies in other species, factors such as IGF and EGF family members, HGF/SF, and various fibroblast growth factors may also be involved with mammary gland development post-puberty [6].

3.4 Pregnancy and Lactation

During pregnancy the mammary gland undergoes ductal morphogenesis, alveolar development, and secretory differentiation in preparation for lactation after parturition. The primary hormones necessary are estrogen, progesterone, prolactin, growth hormone, glucocorticoids, and insulin [27]. Precise hormonal control is not fully known. Estrogen and growth hormone are necessary for ductal morphogenesis. Alveolar development and secretory differentiation require the addition of progesterone and prolactin [51]. During this time genes involved in milk synthesis activate and lactose is excreted at levels positively correlated with increased prolactin levels. Elevated progesterone upregulates TGFβ expression and prevents the active secretion of milk [49]. In many mammals progesterone levels decrease to both

induce partition and begin active secretion of milk. In humans, the placenta is responsible for progesterone synthesis and active secretion does not begin until full removal of the placenta [27].

With the inhibitory effects of progesterone and estrogen decreased, prolactin can induce active secretion to begin, which is about 30–40 h after partition. Four to five days after partition the epithelial cells have fully differentiated to mammary secretory epithelial cells or lactocytes and galactopoiesis, the maintenance of lactation begins. Prolactin and growth hormone are essential for the lactation of the mammary gland [52]. In some species of animals, growth hormone is primarily responsible for lactation but in humans and rodents the influence of prolactin that dominates over lactation. Galactopoietic hormones are not enough to maintain lactation, the continue removal of milk is also necessary. Stimulation of the mammary gland/nipple releases oxytocin from the posterior pituitary and mediates contraction of the smooth muscle found in the myoepithelium surrounding the alveoli of the mammary gland. Contraction of the myoepithelium expels the milk from the lumen of the alveoli into the ducts and out of the gland and allows for continued lactation [27, 31]. Autocrine signaling by a secreted milk protein called feedback inhibitor of lactation also influences milk production. As milk fills the lumen the concentration of feedback inhibitor of lactation increases and it binds to lactocytes inhibiting milk protein trafficking through the Golgi apparatus and milk production decreases. As the mammary gland empties and there is less feedback inhibitor of lactation and milk production increases [53].

3.5 Involution

The hormonal control of involution has been extensively studied in the rodent mammary gland due to its importance in the cancer biology field. The first phase of involution begins with the decrease in prolactin, growth hormone, and epidermal growth factor signaling and an increase in leukemia inhibitory factor. Decreased prolactin activates insulin like growth factor binding protein 5 (IGFBP5) which sequesters IGF and decreases signaling through P-13K resulting in decreased Akt activity and decreased survival signaling [54–56]. Reduced growth hormone signaling also results in decreased IGF levels. Reduction in prolactin, growth hormone, and EGF all result in lowered STAT5 expression and decreased proliferation of the mammary gland [57]. As STAT5 decreases, STAT3 increases due to increased expression of leukemia inhibitory factor [55, 58]. Increased STAT3 induces p50α, p55α, SOCS-3 c-myc, and c/ebpδ to increase pro-apoptotic signals [55, 56]. Increased STAT3 and NF-κB increase inflammation signaling to recruit macrophages. During involution CYP27B1 increases resulting in local conversion of 25-hydroxycholecalciferol to 1α,25-Dihydroxycholecalciferol, which binds to the VDR to directly increase pro-apoptotic signaling, including Caspase8 by increased TNFα signaling [59, 60].

The second phase of involution occurs up to several days after the first phase and is similar to a wound healing. Multiple tissue types are critical for coordinated degradation of the extracellular matrix and extensive remodeling of the mammary lobules. MMP activation leads to the remodel of the extracellular matrix and stroma. MMPs are not activated the first 3 days of involution when TIMP expression is high. As involution progresses, TIMP levels fall and MMPs activate [56, 58]. Increased IGFBP5 inhibits PAI-1 activity which in turn activates t-PA and u-PA to activated plasminogen to plasmin to degrade the basement membrane [61]. Macrophages switch from the pro-inflammatory to an alternatively activated state, often found in other regenerating tissues [62].

4 Chemicals Affecting Mammary Gland Development

4.1 Ethinyl Estradiol

Ethinyl estradiol (EE) is a synthetic, steroidal estrogen used in most oral contraceptives because of its significantly greater oral bioavailability relative to estradiol. EE is also used in contraceptives administered vaginally and transdermally. Currently marketed formulations generally contain between 20 and 35 μg of EE (NTP, 2010). EE has been used for hormone replacement therapy (HRT) at doses between 10 and 50 μg.

EE has been shown to alter mammary gland development in male and female rats after gestational and lactational exposure. Prepubertal development of mammary glands in males and females progressed faster with EE treatment. Females had increased branching and budding and males had an increased number of TEBs [63]. EE exposure during gestation increased mammary cancer risk in several generations of rat offspring due to epigenetic alterations in the mammary gland [64]. The majority of epidemiological studies suggest that there is not a significant increased risk of developing breast cancer from the use of oral contraceptives. Some studies suggest that there is an increased risk of developing breast cancer as the duration of contraceptive use before the first full-term pregnancy increases, and when the start of contraceptives is at a younger age (20–34 years) [65]. In utero exposure to EE resulted in an increased resistance to tamoxifen treatment by induced mammary tumors compared to controls. Resistance to tamoxifen due to in utero exposure to EE may be mediated through reversible epigenetic alterations [66]. HRT is either a combination of estrogen and progestin or estrogen alone. Estrogen and progestin together have been shown to increase the risk of breast cancer and once HRT is stopped, there is an immediate decrease in risk. Estrogen alone is not linked to a higher risk of breast cancer [67].

4.2 Diethylstilbestrol

Diethylstilbestrol (DES) was the first synthetic nonsteroidal estrogen. The trans-isomer is used for commercial purposes and is stable in the environment [68]. It was widely prescribed in the USA from

the early 1940s until 1971, to prevent miscarriages and premature deliveries. Other uses included treatment of symptoms arising from menopause, senile vaginitis, postcoital, emergency contraceptive, prostate, and breast cancer [69]. Up to 10 million people in the USA received DES or were exposed in utero between 1940 and 1970. DES is an endocrine disrupting compound, a transplacental carcinogen, and a teratogen able to induce developmental defects. DES is classified as a carcinogen by the US EPA, NTP, and IARC and is no longer commercially available in the USA [69].

Women, who were prescribed DES during pregnancy, had an increase in the risk of invasive breast cancer [69]. The daughters of women who took DES during pregnancy had an increased risk of developing clear cell adenocarcinoma of the vagina and cervix [69]. DES exposure in utero also increases the breast cancer risk for women at 40 years of age or older [70].There is a positive association between men exposed to DES in utero and an increased risk of testicular cancer. DES is also known to affect endocrine sensitive tissues and may have hereditary effects due to DNA modifications [71].

In rodent studies, neonatal exposure to DES exerts long lasting effects on the mammary gland. Mice given 50 μg of neonatal DES had dilatation of the mammary ducts and precocious lactogenesis [72]. DES exposure has caused mammary gland cancer both (benign and malignant) in rodents when treated at adults. The mammary gland has a distinct biphasic response to DES exposure in both female and male mice. DES at lower concentrations stimulated mammary growth, but higher concentrations inhibited mammary gland development in male and female mice [72].

4.3 Genistein

Genistein is a naturally occurring isoflavone, part of a group of estrogenic plant-based compounds called phytoestrogens. Genistein is one of the most abundant and bioactive compounds in soy. Most human exposure comes from the consumption of food, infant formulas, and/or dietary supplements made with soybeans and soy protein [73]. Soy-based infant formula is of special concern due the fact that infants fed soy formula were demonstrated to have plasma isoflavone blood levels exceeding those of Japanese adults severalfold and due to the potentially sensitive window of exposure [74]. Genistein has also been found in breast milk and can cross the placenta.

Both beneficial and adverse effects have been reported due to genistein exposure [75]. The timing and the dose of genistein may be indicative of whether a protective or adverse effect occurs. Perinatal and prepubertal genistein exposure in female rats has been shown to reduce the risk of developing carcinogen-induced mammary tumors [76, 77]. Potential modes of action include increasing differentiation of the mammary epithelium or by upregulating the tumor suppressing gene BRCA1 in the mammary gland [78–80]. A reduction of TEBs reduces the sites of susceptibility to carcinogens. Prenatal genistein exposure increased the number of TEBs and induced

ductal hyperplasias in rats and mice [81–83]. Male rat mammary glands are more sensitive to endocrine disruption by genistein [81]. In a multi-generational genistein and ethinyl estradiol studies male rats developed mammary gland hyperplasia, but not feminization [84]. In cancer bio-assays, some or equivocal evidence were found of carcinogenic activity in female rats [73]. Epidemiology studies on adolescent exposure to soy products in the diet align with the animal studies, soy components provide protection against breast cancer later in life. Gestational genistein exposure and breast cancer risk have not been evaluated [75].

4.4 Bisphenol A

Bisphenol A (BPA) is a synthetic organic compound used primarily in the production of plastics and epoxy resins. BPA is one of the highest volume chemicals produced. In 2011, worldwide production was estimated at 10 billion pounds [85]. Many consumer products contain BPA, such as polycarbonate bottles, the lining in metal food and drink cans, dental sealants, and receipt paper. BPA has a short half-life in adults (5.3 h in humans and 90 min in rats) but exposure may be consistent due to the many different sources of exposure many people experience [86]. Humans are primarily exposed to BPA through the diet, though inhalation of dust or skin exposure through water is possible. Infants and children typically have the highest intake of BPA (excluding occupational exposure) due to eating more comparability by weight and spending more time on the floor compared to adult [87].

BPA is a weakly estrogenic chemical and can cause endocrine disruption. BPA can bind to both ERα and β. In rodent studies, low dose, prenatal exposure to BPA, administered via subcutaneous mini-pump, caused changes in the mammary gland that may indicate an increased susceptibility to develop mammary gland tumors [87]. Some of the changes in the mammary gland found in these studies include an altered rate of maturation, increased ductal growth, formation of intra-ductal hyperplasia (beaded ducts), and increased early TEB formation [88–91]. Suggesting prenatal BPA effects breast tissue maturation that may lead to a predisposition to disease onset later in life. No human study has examined BPA exposure in early life and adult breast cancer risk. Based on the current data there is a minimal concern for the pre- and peri-natal low dose effects on BPA on the mammary gland [13, 92–95].

The NTP has determined that there is minimal concern for the effects of BPA exposure on the mammary gland for females in fetuses, infants, and children [95]. Despite current perspective of the FDA and European Food Safety Authority (EFSA) that BPA is safe at the current levels occurring in foods [96–98], BPA is being voluntarily withdrawn from many consumer products due to pressure from the consumers. The US EPA and EFSA set the reference dose/tolerable daily intake at 50 and 4 μg/kg, respectively [96, 97]. A growing concern is the safety of “BPA-free” alternatives currently used to replace BPA, such as Bisphenol S.

4.5 Triclosan

Triclosan (TCS) is an antibacterial and antifungal agent found in consumer products, including toothpaste, soaps, detergents, toys, and surgical cleaning treatments. Human exposure to triclosan occurs primarily through use of personal care products, such as toothpastes, deodorants, and soaps [99].

TSC is marketed under a variety of trademarked names including Microban, Irgasan DP-300, Lexol 300, Ster-Zac, Cloxifenolum, Biofresh, and others its annual usage in 2005 was estimated to be over 300 tons in the USA and over 350 tons in Europe [100]. TCS does not persist in the body following absorptions as free TSC and its conjugates are rapidly eliminated within 24 h [101].

Concerns for the possible effects on human health have been growing over the past decade following the detection of TSC in human plasma, milk, and urine with levels of TCS in the blood correlating with consumer use patterns of the antimicrobial. TSC was found in breast milk samples, in the bile of fish exposed to municipal wastewater and in wild living fish from the receiving waters of the three wastewater treatment plants in Sweden [102]. Likewise, TCS was also detected in the urine of the US population [103]. Plasma concentration was observed to be higher than in milk in mothers exposed to consumer products containing TSC, indicating that infant exposure to triclosan via breast milk is much less than the dose in the mother [104].

Studies in various species provide strong evidence that TSC is an endocrine disruptor. TCS alters serum thyroid hormone and testosterone concentrations in male rats [105]. It is reported here that triclosan possesses intrinsic estrogenic and androgenic activity in a range of assays in vitro [106]. Moreover, TCS affects estrogen-mediated response in the pubertal and weanling female rat and also suppressed thyroid hormone [107].

The established role of estrogen in the development and progression of breast cancer raises questions concerning a potential contribution from the many chemicals in the environment which can enter the human breast and which have estrogenic activity. TSC possess estrogen-mimicking properties and has been measured in human breast adipose tissue and in human milk. This raises concerns regarding its association with various health outcomes, including breast cancer development [108]. Since human studies are lacking, the need for epidemiologic studies of risk associated with various concentrations and durations of exposure to TSC has been suggested, as well as studies to characterize human exposure to the varying use of triclosan-containing consumer products and other routes of exposure [109].

Due to concerns that TSC (liquid soaps) and triclocarban (bar soaps) could pose health risks, such as bacterial resistance or hormonal effects, the US FDA has issued a proposed rule that requires manufacturers of antibacterial hand soaps and body washes to demonstrate that their products are safe for long-term daily use and more effective than plain soap and water in preventing illness and

the spread of certain infections [110]. This proposed rule does not affect hand sanitizers, wipes, or antibacterial products used in health care settings. Just few months ago and exactly in September, 2016, the US FDA has issued a final rule establishing that over-the-counter (OTC) consumer antiseptic wash products containing the most commonly used ingredients, triclosan and triclocarban can no longer be marketed [111] based on some data suggesting that antibacterial ingredients may do more harm than good over the long-term.

4.6 Vinclozolin

Vinclozolin is a dicarboximide pesticide that is used for the control of several species of fungi in vines (such as grapes), strawberries, vegetables, fruit, and ornamentals. Exposure of vinclozolin to human comes mainly from the consumption of residual contamination in foods and drinking water. The compound is a lipophilic chemical capable of bioaccumulating in fat tissues [112].

There is evidence that vinclozolin binds fairly weakly to the androgen receptor but that at least two vinclozolin metabolites occurring in mammals, plants, and soil are responsible for much of the antiandrogenic activity attributable to vinclozolin [113, 114]. Two active metabolites of vinclozolin, M1 and M2, were found to compete with natural androgens, an effect that blocks androgen-induced gene expression in vivo and in vitro following observation that low-level exposure to these metabolites may be associated with reproductive toxicity of vinclozolin, namely nipple retention, reduced ano-genital distance (AGD), and reduced seminal vesicle and ventral prostate weights [113, 115, 116]. Female mice with the androgen receptor knocked out have impaired mammary ductal growth in postnatal life including altered ductal elongation [117].

Moreover, there is growing evidence identified between EDCs altering the epigenomic landscape in cancers, including breast, and common diseases like cardiovascular, pulmonary, or neurodegenerative disorders [118]. In such cases, early life epigenetic programming and also those encountered during adult life are believed to be largely contributing factors to overall progression of disease states.

Exposure of gestating females to environmental factors or toxicants during the period of gonadal sex determination can alter epigenetic transgenerational inheritance. Transient exposure of gestating inbred and CD1 outbred mice on GD 7–13 to vinclozolin induced transgenerational disease in the outbred CD-1 strain, but not the inbred strain [119]. Likewise vinclozolin has been shown to promote an epigenetic alteration in the male rat germ line that appears to transmit a transgenerational disease state [120]. Analysis of the F3 generation sperm epigenome identified differential DNA methylation regions [119, 121]. Although vinclozolin is not classified as mammary carcinogen in human, but it was found to induce a low incidence of mammary tumors in female offspring following a multigenerational study in rats [121] and disrupts mammary gland development, where epithelial

branching and terminal end buds (TEBs) are increased compared with controls [122, 123].

This suggests a potential epigenetic etiology and molecular basis of adult-onset disease. The development of environmental epigenetic biomarkers may be more suitable for the prediction of future disease risk, including that for breast cancer [124, 125].

4.7 2,3,7,8-Tetrachlorodibenzo-p-Dioxin

2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) is a persistent polychlorinated dibenzodioxin side product primarily in organic synthesis and burning of organic materials. Although it has never been produced commercially it is formed as a side product when producing certain herbicides, most famously in Agent Orange. Ingestion of contaminated food is the primary source of dioxin exposure and can accumulate in fatty tissues. Exposed mothers can transfer TCDD both to the fetus in utero and to the infant via the breast milk [126].

In countries with less pollution and stricter emission standards in effect, the body burden in people is approximately 2 ppt. In many developing nations the body burdens can be significantly higher. TCDD is chemically stable and not readily metabolized in most species and thus exhibits a significant degree of bioaccumulation and environmental persistence [127].

TCDD is the most potent aryl hydrocarbon receptor (AhR) agonist. When bound to TCDD, AhR disassociates with chaperon proteins and dimerizes with AhR nuclear translocator (ARNT). AhR/ARNT is transported to the nucleus where AhR induces gene transcription of a variety of pleiotropic proteins [128]. TCDD exhibits a broad spectrum of antiestrogenic responses through AhR activation. ER in the mammary gland is downregulated [129]. In animal studies, TCDD administered during critical windows of susceptibility affects mammary gland development in both males and females. When exposed prenatally, developmental delays were seen starting at birth and persisted through to adulthood [130]. After gestational and lactation exposure the mammary glands shortly after parturition have reduced primary branches, decreased elongation, and fewer alveolar buds and lateral branching [131]. During puberty, the gestational exposed animals have delayed mammary gland development with stunted progression of the epithelium and delayed TEB differentiation. At early adulthood, TCDD-exposed glands had many undifferentiated ducts, fewer branches and in females, often failed to full the fat pad. TCDD also severely impairs mammary gland differentiation when mice were exposed during pregnancy, resulting in stunted growth, poor lobuloalveolar differentiation, and suppressed whey acidic protein expression in the lactating mammary gland [132, 133].

TCDD is a known multisite human carcinogen [134]. In rats, prenatal TCDD exposure increased susceptibility to chemically induced mammary adenocarcinomas. Epidemiology studies have demonstrated that serum levels of TCDD are positively associated with

a delay in breast development and an increased breast cancer risk [135, 136].

4.8 Cadmium

Cadmium (Cd) is a ubiquitous environmental pollutant, widely dispersed through industrial emission, waste incineration, and combustion of fossil fuels. The variety of sources of exposure includes occupational exposure from mines, metal smelters, and other industries which use Cd compounds. The most common means of exposure is from smoking tobacco and exposure to second hand smoke. All forms of tobacco contain high levels of Cd. The main sources of dietary cadmium exposure are bread and other cereals, potatoes, root crops, and vegetables [137]. Cd accumulates in the body, primarily in the liver and kidneys. The estimated total body burden is 9.5–40 mg in the USA and Europe [138].

Cd is a known human carcinogen. High levels of Cd in the urine has been linked to an increased risk of developing breast cancer in women. Cd is an EDC and has been shown to mimic the effects of estrogen, by binding to ERα [139]. Cd induced gene expression similar to estradiol-induced gene expression in in vitro human mammary epithelial cells and promoted both uterine and mammary gland growth in vivo rodent models [139, 140]. In utero exposure increased both the number of TEBs and pre-neoplastic lesions in rats later in life [141]. Whereas, in prepuberal mice, Cd exposure disturbed mammary ductal growth and reduced the number of TEBs. In adult mice, Cd again mimicked estrogen by increasing lobuloalveolar development and ductal branching. In the lactating mammary gland in mice, Cd exposure increased the fat content in the mammary gland, condensing the appearance of the alveoli, decreased β-casein gene expression, and reduced levels of calcium in the mammary glands [142]. These effects on lactation may impair the development of offspring. Several epidemiological studies suggest a link between occupational exposure to Cd and an increased risk of breast cancer [143, 144].

4.9 Perfluorooctanoic Acid

Perfluorooctanoic acid (PFOA) is an eight-carbon perfluoroalkyl acid, a member of a group of chemicals called perfluoroalkyl and polyfluoroalkyl substances (PFAS). It is produced both synthetically and through the degradation of other PFAS. PFOA is commonly used as a water and oil repellent for fabric coatings, in food storage containers, lubricants, and fire extinguishing foams [145]. Although PFOA production is being phased out in the USA, it is persistent in the environment. PFOA has a long half-live in humans, between 3.8 and 4.4 years, and is found ubiquitously in human serum [146].

In animal studies, adult exposures in rats resulted in hepatomegaly and a common triad of tumors, hepatocellular adenoma, Leydig cell tumors, and pancreatic acinar cell tumors. This tumor triad is believed to be mediated by peroxisome proliferator-activated receptor-alpha (PPARα), a non-human relevant mechanism [147].

Developmental exposure to PFOA results in a variety of toxicities in mice including prenatal loss, reduced birth weight, delayed eye opening, increased postnatal mortality, and delayed sexual maturation in females and is also believed to be dependent on PPARα [147]. There are strain differences in sensitivity to low doses of PFOA in mice. In one strain of mice, peripubertal PFOA exposure delayed mammary gland development, while had mixed effects in a different mouse strain [148]. For mice strains sensitive to developmental PFOA exposure, mammary gland delays occurred in the absence of liver effects [149–151]. PPARα may not be required for alterations in mammary gland development. PFOA treatment has been shown to affect mammary gland development in PPARα knockout mice [152]. A mode of action for a non-PPARα mediated mode of action is not fully understood. PFOA does not activate or bind to ER but did enhance the estrogenic effects of estradiol in T47D human breast cancer cells. Also men with higher PFASs in their serum had higher estradiol levels as well. Epidemiology data on the effects of PFOA exposure are limited and often conflicting. There has been evidence that exposure to PFAS is associated with a longer pregnancy time and abnormal birth outcomes. In addition, PFAS, including PFOA, may contribute to the risk of breast cancer, found in a small cohort of Greenlandic Inuit women [153]. The EPA and IARC have determined that PFOA has "suggestive evidence of carcinogenic potential" and "possibly carcinogenic to humans" [154, 155].

4.10 Atrazine

Atrazine is a selective pre- and post-emergence herbicide for annual control of grass and broad-leaved weeds in many different crops and as a soil sterilant for airfields. It is the second most commonly used herbicide in the USA. Atrazine is highly water soluble, has a long half-life and water contamination is widespread. There are many different trade names for atrazine and many other herbicides may be formulated with atrazine. An estimated 31–35 million kg of atrazine was used annually in the USA for the past 35 years [156]. Use of atrazine was banned in the European Union since 2004.

The importance of atrazine as an EDC is primary due to its effects on the hypothalamic–pituitary–gonadal (HPG) axis which is vital for maintaining proper reproductive function [157]. Effects on the mammary gland have been shown predominately in rats. Prenatal exposure to atrazine has been shown to alter mammary gland development in rats, resulting in glands with delayed TEB presence, sparse branching patterns, and impaired growth [158, 159]. Atrazine has been reported to reduce GnRH pulses and decreased LH pulse frequency while increasing pulse period and amplitude. Suppression on LH causes altered PRL regulation and disrupts the normal estrous cycle in rats [157]. Mammary tumors have been shown to occur in some rat stains, but not all. In sensitive rat strains atrazine exposure leads to premature reproductive senescence [160]. In the post-reproductive rat, estrogen/prolactin stimulates the mammary

gland and increases mammary tumor formation, while in post-menopausal women, there is minimal estrogen/prolactin stimulation. This mode of action is believed to not to be a human relevant mechanism [157, 161].

Epidemiological studies on atrazine (or its metabolites), which often have limitations, have not found an association between atrazine and breast cancer and the EPA and IARC have determined there is inadequate evidence in humans for the carcinogenicity of atrazine [156, 162, 163]. TEBs are target sites for chemical carcinogens. The extended presence of TEBs of mammary glands in prenatally exposed rats suggests that gestational exposure to atrazine could increase the window of susceptibility of the mammary gland in humans and could be the first hit in a "two hit" model of breast cancer [164].

4.11 Radiation

Both epidemiology and animal studies support the direct relationship of increased radiation exposure and increased breast cancer risk. Radiation exposure during adolescent and early adulthood (under 20 years old) has been shown to particularly increase the risk of breast cancer development [165]. Besides causing both direct and indirect DNA damage that leads to tumorigenesis, the mammary gland is also susceptible to low levels of radiation that affect the cell signaling [143, 165]. Irradiation of just the mammary stroma and not the mammary epithelium has been shown to affect the microenvironment of the mammary gland causing accelerated development of tumors [166]. Irradiation alters the microenvironment to activate TGFβ, induces extracellular matrix remodeling, affects cell fate decisions, mediates ATM kinase control of the DNA damage response, and deregulates mammary stem cells [166].

5 Implication of Altered Development

Exposure to environmental chemicals may lead to perturbations in mammary gland (MG) development and increase the risk for later adverse effects, including lactation impairment, gynecomastia (in males), and breast [8, 167–169]. Early childhood exposure to EDCs, including atrazine, bisphenol A (BPA), dibutylphthalate, TCDD, methoxychlor, found to induce altered MG development namely, ductal hyperplasia, alveolar hypoplasia, delayed MG development, reduced apoptosis in TEBs, and increased or decreased numbers of terminal ducts or lobules [87, 170–172].

After weaning the mammary gland undergoes post-lactational and lobular involutions. During the post-lactational involution the mammary gland undergoes massive cell death and tissue remodeling as it begins to return to the pre-pregnant state. Whereas, lobular involution is the process by which the breast epithelial tissue is gradually lost with aging of the mammary gland. Although post-

lactational and lobular involutions are distinct processes, studies have indicated that both are related to breast cancer development [173]. MG development is a complex process that extends from gestation through adolescent; therefore, chemical exposure during susceptible windows of development may alter the MG and can be the risk of breast cancer, as well as other adverse outcomes [167].

Epidemiological studies provide no support for a causal relationship between atrazine exposure and breast cancer. This conclusion is consistent with International Agency for Research on Cancer's (IARC) classification of atrazine as "unclassifiable as to carcinogenicity" and the USEP's classification of atrazine as "not likely to be carcinogenic" [174]. In contrast TCDD is a known human carcinogen causing an increase in total cancer [175] based on limited evidence in humans, sufficient evidence in animals, and extensive mechanistic information indicating that TCDD acts through a mechanism involving the aryl hydrocarbon receptor (AhR), which is present in both humans and animals. Steenland et al. presented new evidence that supported the 1997 IARC classification [176]. Later epidemiological studies fall far short of conclusively demonstrating a causal link between TCDD exposure and cancer risk in humans [177]. Likewise, there is a conflicting human data on the role of TCDD on breast cancer; although a Seveso (Italy) study showed a significant increase in breast cancer on women with higher TCDD levels [135]. However, later studies were inconclusive.

The health consequences of diethylstilbestrol (DES) in later generation have been widely reported. Herbst et al. (1971) established DES as a human carcinogen in a study linking cases of the rare vaginal clear cell carcinoma to daughters of mothers treated with DES during pregnancy [178]. Similarly, a number of animal studies have been performed to further confirm the link between DES exposure and breast cancer. Although there is limited evidence to support an epigenetic basis for DES-induced breast cancer risk, nevertheless, increasing evidence has implicated a role for epigenetics in DES-related mammary tumor [179, 180].

Altered MG is also associated with non-cancer effects such as impaired lactation and gynecomastia. The hormonal control of lactation is entirely similar across mammalian species, but it is really questionable the utility of the rodent as a model for predicting chemical effects on human lactation. Gynecomastia, an altered MG development which occurs in up to 2/3 of pubertal boys is believed to be due to an imbalance in estrogen and testosterone [181]. However, there is limited data to support the association between gynecomastia and EDCs. Nevertheless, a number of epidemiological studies of prepubertal gynecomastia suggest an association with exposure to estrogenic or androgenic substances in boys [182, 183].

Epidemiological studies of breast cancer risk are generally consistent with rodent bioassays, although only a few agents have been studied in humans. The rodent models are useful as screening tools

for identifying potential breast carcinogens, an expert panel on MG tumors concluded [184]. Interestingly, most chemicals that are positive for breast tumors in the rodent cancer bioassay are genotoxic and many are considered multisite carcinogens, supporting their relevance to humans [185]. Endocrine disruptors are not generally genotoxicants, yet they can induce significant adverse outcome.

6 Concluding Remarks

Mammary gland development is a complex process involving postnatal maturation. Environmental and lifestyle factors are considered to among the major influencing components increasing breast cancer risk in human, namely age at menopause, age at menarche, parity, and a women's age at her first pregnancy. Early exposure to environmental chemicals may have deleterious effects later in life including the development of breast cancer. Association between breast cancer and DDT exposure was demonstrated only when exposure occurred before 14 years of age [186]. Currently, there are only limited evidence of how chemicals such as endocrine disruptors may alter mammary gland development. Likewise, there is no conclusive evidence that EDCs induce breast cancer in humans, but the correlation of male and female breast cancer incidence rates suggests the existence of risk factors including environmental chemicals [187]. On the other hand, it has been demonstrated that early exposure to EDCs induces epigenetic transgenerational disease and hence promotes an adult-onset disease.

Overall, more research studies are required to advance the etiology of breast cancer and the EDCs. Enhancing the understanding of breast cancer etiology as well as non-cancer effects such as altered lactation and gynecomastia requires elucidating the mechanisms involved in breast cancer development. In addition, the elucidation of the role of epigenetics in environment–genome interactions will provide critical insights for environmental health and disease [188].

References

1. Brisken C, Park S, Vass T, Lydon JP, O'Malley BW, Weinberg RA (1998) A paracrine role for the epithelial progesterone receptor in mammary gland development. Proc Natl Acad Sci U S A 95(9):5076–5081
2. Hens JR, Wysolmerski JJ (2005) Key stages of mammary gland development: molecular mechanisms involved in the formation of the embryonic mammary gland. Breast Cancer Res 7 (5):220–224
3. Maller O, Martinson H, Schedin P (2010) Extracellular matrix composition reveals complex and dynamic stromal-epithelial interactions in the mammary gland. J Mammary Gland Biol Neoplasia 15(3):301–318
4. Brisken C, O'Malley B (2010) Hormone action in the mammary gland. Cold Spring Harb Perspect Biol 2(12):a003178
5. Hennighausen L, Robinson GW (2001) Signaling pathways in mammary gland development. Dev Cell 1(4):467–475
6. Hovey RC, Trott JF, Vonderhaar BK (2002) Establishing a framework for the functional mammary gland: from endocrinology to

morphology. J Mammary Gland Biol Neoplasia 7(1):17–38

7. Howard BA, Gusterson BA (2000) Human breast development. J Mammary Gland Biol Neoplasia 5(2):119–137
8. Rudel RA, Fenton SE, Ackerman JM, Euling SY, Makris SL (2011) Environmental exposures and mammary gland development: state of the science, public health implications, and research recommendations. Environ Health Perspect 119(8):1053–1061
9. Howlader N, Noone AM, Krapcho M, Miller D, Bishop K, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds) (2016) SEER cancer statistics review, 1975–2013. National Cancer Institute, Bethesda, MD. http://seer.cancer.gov/csr/1975_2013/ (based on November 2015 SEER data submission)
10. American Cancer Society (ACS) (2016) Cancer facts and figures 2016. American Cancer Society, Atlanta
11. American Cancer Society (ACS) (2015) Breast cancer facts and figures 2015–2016. American Cancer Society, Inc., Atlanta
12. IBCERCC (2013) www.niehs.nih.gov/about/boards/ibcercc
13. Macon MB, Fenton SE (2013) Endocrine disruptors and the breast: early life effects and later life disease. J Mammary Gland Biol Neoplasia 18(1):43–61
14. Forman MR, Winn DM, Collman GW, Rizzo J, Birnbaum LS (2015) Environmental exposures, breast development and cancer risk: through the looking glass of breast cancer prevention. Reprod Toxicol 54:6–10
15. Drife JO (1986) Breast development in puberty. Ann N Y Acad Sci 464(1 Endocrinology):58–65
16. Russo J, Russo IH (2004) Development of the human breast. Maturitas 49(1):2–15
17. Jolicoeur F (2005) Intrauterine breast development and the mammary myoepithelial lineage. J Mammary Gland Biol Neoplasia 10(3):199–210
18. Robinson GW, Karpf ABC, Kratochwil K (1999) Regulation of mammary gland development by tissue interaction. J Mammary Gland Biol Neoplasia 4(1):9–19
19. Macias H, Hinck L (2012) Mammary gland development. Wiley Interdiscip Rev Dev Biol 1(4):533
20. Gusterson BA, Stein T (2012) Human breast development. Semin Cell Dev Biol 23(5):567
21. Hassiotou F, Geddes D (2013) Anatomy of the human mammary gland: current status of knowledge. Clin Anat 26(1):29–48
22. Javed A, Lteif A (2013) Development of the human breast. Semin Plast Surg 27(1):005–012
23. Alexander JM, Campbell MJ (1997) Prevalence of inverted and non-protractile nipples in antenatal women who intend to breastfeed. Breast 6(2):72–78
24. Montagna W, Yun JS (1972) The glands of montgomery. Br J Dermatol 86(2):126–133
25. Anbazhagan R, Bartek J, Monaghan P, Gusterson BA (1991) Growth and development of the human infant breast. Am J Anat 192(4):407–417
26. Marshall WA, Tanner JM (1969) Variations in pattern of pubertal changes in girls. Arch Dis Child 44(235):291–303
27. Neville MC, McFadden TB, Forsyth I (2002) Hormonal regulation of mammary differentiation and milk secretion. J Mammary Gland Biol Neoplasia 7(1):49–66
28. Anderson E, Clarke RB (2004) Steroid receptors and cell cycle in normal mammary epithelium. J Mammary Gland Biol Neoplasia 9(1):3–13
29. Visvader JE (2009) Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. Genes Dev 23(22):2563–2577
30. Russo J, Lynch H, Russo IH (2001) Mammary gland architecture as a determining factor in the susceptibility of the human breast to cancer. Breast J 7(5):278–291
31. Beesley R, Johnson J (2008) Glob. libr. women's med
32. Davis B, Fenton S (2013) Mammary gland. In: Haschek WM, Rousseaux CG, Wallig MA (eds) Haschek and Rousseaux's handbook of toxicologic pathology, vol 3. Elsevier Academic Press, New York, NY, pp 2665–2694
33. Lamote I, Meyer E, Massart-Leën AM, Burvenich C (2004) Sex steroids and growth factors in the regulation of mammary gland proliferation, differentiation, and involution, vol 69. Elsevier Inc., United States
34. Taylor D, Pearce CL, Hovanessian-Larsen L, Downey S, Spicer DV, Bartow S et al (2009) Progesterone and estrogen receptors in pregnant and premenopausal non-pregnant normal human breast. Breast Cancer Res Treat 118(1):161–168
35. Russo J, Russo IH (2014) Techniques and methodological approaches in breast cancer research, vol 1. Springer, New York, NY

36. Cowin P, Wysolmerski J (2010) Molecular mechanisms guiding embryonic mammary gland development. Cold Spring Harb Perspect Biol 2(6):a003251–a003251
37. Robinson GW, Hennighausen L (1997) Inhibins and activins regulate mammary epithelial cell differentiation through mesenchymal-epithelial interactions. Development 124 (14):2701
38. Hatsell S, Rowlands T, Hiremath M, Cowin P (2003) β-catenin and tcfs in mammary development and cancer. J Mammary Gland Biol Neoplasia 8(2):145–158
39. Lindvall C, Bu W, Williams BO, Li Y (2007) Wnt signaling, stem cells, and the cellular origin of breast cancer. Stem Cell Rev 3 (2):157–168
40. Osin PP, Anbazhagan R, Bartkova J, Nathan B, Gusterson BA (1998) Breast development gives insights into breast disease. Histopathology 33(3):275–283
41. Gallego MI, Binart N, Robinson GW, Okagaki R, Coschigano KT, Perry J, Kopchick JJ, Oka T, Kelly PA, Hennighausen L (2001) Prolactin, growth hormone, and epidermal growth factor activate Stat5 in different compartments of mammary tissue and exert different and overlapping developmental effects. Dev Biol 229(1):163–175
42. Feldman M, Ruan W, Tappin I, Wieczorek R, Kleinberg D (1999) The effect of GH on estrogen receptor expression in the rat mammary gland. J Endocrinol 163(3):515–522
43. Ciarloni L, Mallepell S, Brisken C (2007) Amphiregulin is an essential mediator of estrogen receptor α function in mammary gland development. Proc Natl Acad Sci U S A 104(13):5455–5460
44. McBryan J, Howlin J, Napoletano S, Martin F (2008) Amphiregulin: role in mammary gland development and breast cancer. J Mammary Gland Biol Neoplasia 13(2):159–169
45. Obr AE, Edwards DP (2012) The biology of progesterone receptor in the normal mammary gland and in breast cancer. Mol Cell Endocrinol 357(1–2):4–17
46. Aupperlee MD, Leipprandt JR, Bennett JM, Schwartz RC, Haslam SZ (2013) Amphiregulin mediates progesterone-induced mammary ductal development during puberty. Breast Cancer Res 15(3):R44
47. Beleut M, Rajaram RD, Caikovski M, Ayyanan A, Germano D, Choi Y et al (2010) Two distinct mechanisms underlie progesterone-induced proliferation in the mammary gland. Proc Natl Acad Sci U S A 107(7):2989–2994
48. Forsyth IA, Wallis M (2002) Growth hormone and prolactin – molecular and functional evolution. J Mammary Gland Biol Neoplasia 7(3):291–312
49. Monks J (2007) TGFβ as a potential mediator of progesterone action in the mammary gland of pregnancy. J Mammary Gland Biol Neoplasia 12(4):249–257
50. Navarrete MAH, Maier CM, Falzoni R, de Azevedo Quadros LG, Lima GR, Baracat EC, Nazário ACP (2005) Assessment of the proliferative, apoptotic and cellular renovation indices of the human mammary epithelium during the follicular and luteal phases of the menstrual cycle. Breast Cancer Res 7(3): R306–R313
51. Oakes SR, Rogers RL, Naylor MJ, Ormandy CJ (2008) Prolactin regulation of mammary gland development. J Mammary Gland Biol Neoplasia 13(1):13–28
52. Harris J, Stanford PM, Sutherland K, Oakes SR, Naylor MJ, Robertson FG, Blazek KD, Kazlauskas M, Hilton HN, Wittlin S, Alexander WS, Lindeman GJ, Visvader JE, Ormandy CJ (2006) Socs2 and Elf5 mediate prolactin-induced mammary gland development. Mol Endocrinol 20(5):1177–1187
53. Knight CH, Peaker M, Wilde CJ (1998) Local control of mammary development and function. Rev Reprod 3(2):104–112. doi:10.1530/revreprod/3.2.104
54. Allan GJ, Beattie J, Flint DJ (2004) The role of IGFBP-5 in mammary gland development and involution. Domest Anim Endocrinol 27 (3):257–266
55. Baxter FO, Neoh K, Tevendale MC (2007) The beginning of the end: death signaling in early involution. J Mammary Gland Biol Neoplasia 12(1):3–13
56. Sutherland KD, Lindeman GJ, Visvader JE (2007) The molecular culprits underlying precocious mammary gland involution. J Mammary Gland Biol Neoplasia 12(1):15–23
57. Barash I (2006) Stat5 in the mammary gland: controlling normal development and cancer. J Cell Physiol 209(2):305–313
58. Watson C (2006) Key stages in mammary gland development – involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ. Breast Cancer Res 8(2):203
59. Furth PA (1999) Mammary gland involution and apoptosis of mammary epithelial cells. J Mammary Gland Biol Neoplasia 4 (2):123–127
60. Zinser GM, Welsh J (2004) Accelerated mammary gland development during pregnancy

and delayed postlactational involution in vitamin D3 receptor null mice. Mol Endocrinol 18(9):2208–2223

61. Tonner E, Allan G, Flint D (2000) Hormonal control of plasmin and tissue-type plasminogen activator activity in rat milk during involution of the mammary gland. J Endocrinol 167(2):265–273
62. Hughes K, Wickenden JA, Allen JE, Watson CJ (2012) Conditional deletion of Stat3 in mammary epithelium impairs the acute phase response and modulates immune cell numbers during post-lactational regression. J Pathol 227(1):106–117
63. Mandrup KR, Hass U, Christiansen S, Boberg J (2012) Perinatal ethinyl oestradiol alters mammary gland development in male and female Wistar rats. Int J Androl 35 (3):385–396
64. de Assis S, Warri AM, Cruz I, Laja O, Tian Y, Zhang B, Wang Y, Huang TH, Hilakivi-Clarke L (2012) High-fat or ethinyl-oestradiol intake during pregnancy increases mammary cancer risk in several generations of offspring. Nat Commun 3:1053
65. Wingo PA, Lee NC, Ory HW, Beral V, Peterson HB, Rhodes P (1993) Age-specific differences in the relationship between oral contraceptive use and breast cancer. Cancer 71(S4):1506–1517
66. Hilakivi-Clarke L, Wärri A, Bouker KB, Zhang X, Cook KL, Jin L, Zwart A, Nguyen N, Hu R, Cruz MI, de Assis S, Wang X, Xuan J, Wang Y, Wehrenberg B, Clarke R (2017) Effects of in utero exposure to ethinyl estradiol on tamoxifen resistance and breast cancer recurrence in a preclinical model. J Natl Cancer Inst 109(1.) p.djw188
67. American Cancer Society (ACS) (2015) Menopausal hormone therapy and cancer risk. American Cancer Society, Inc., Atlanta
68. NTP National Toxicology Program (2011) Diethylstilbestrol. 12th Report on Carcinogens 12:159–161
69. IARC (2012) Diethylstilbestrol. A review of human carcinogens. IARC Monogr Eval Carcinog Risks Hum 100A:175–218
70. Reed CE, Fenton SE (2013) Exposure to diethylstilbestrol during sensitive life stages: a legacy of heritable health effects. Birth Defects Res C Embryo Today 99(2):134
71. Palmer JR, Wise LA, Hatch EE, Troisi R, Titus-Ernstoff L, Strohsnitter W, Kaufman R, Herbst AL, Noller KL, Hyer M, Hoover RN (2006) Prenatal diethylstilbestrol exposure and risk of breast cancer. Cancer Epidemiol Biomarkers Prev 15(8):1509–1514
72. Hovey RC, Asai-Sato M, Warri A, Terry-Koroma B, Colyn N, Ginsburg E, Vonderhaar BK (2005) Effects of neonatal exposure to diethylstilbestrol, tamoxifen, and toremifene on the BALB/c mouse mammary gland. Biol Reprod 72(2):423–435
73. NTP National Toxicology Program (2008) Toxicology and carcinogenesis studies of genistein (cas no. 446-72-0) in Sprague-Dawley rats (feed study). National Toxicology Program Technical Report Series, vol (545), p 1
74. Strom BL, Schinnar R, Ziegler EE, Barnhart KT, Sammel MD, Macones GA, Stallings VA, Drulis JM, Nelson SE, Hanson SA (2001) Exposure to soy-based formula in infancy and endocrinological and reproductive outcomes in young adulthood. JAMA 286 (7):807–814
75. Spagnuolo C, Russo GL, Orhan IE, Habtemariam S, Daglia M, Sureda A, Nabavi SF, Devi KP, Loizzo MR, Tundis R, Nabavi SM (2015) Genistein and cancer: current status, challenges, and future directions. Adv Nutr 6 (4):408
76. Lamartiniere CA, Zhang JX, Cotroneo MS (1998) Genistein studies in rats: potential for breast cancer prevention and reproductive and developmental toxicity. American Society for Clinical Nutrition, Inc., United States
77. Hilakivi-Clarke L, Onojafe I, Raygada M, Cho E, Skaar T, Russo I, Clarke R (1999) Prepubertal exposure to zearalenone or genistein reduces mammary tumorigenesis. Br J Cancer 80(11):1682–1688
78. Delclos KB, Bucci TJ, Lomax LG, Latendresse JR, Warbritton A, Weis CC, Newbold RR (2001) Effects of dietary genistein exposure during development on male and female CD (Sprague-Dawley) rats. Reprod Toxicol 15(6):647–663
79. Cabanes A, Wang M, Olivo S, DeAssis S, Gustafsson J, Khan G, Hilakivi-Clarke L (2004) Prepubertal estradiol and genistein exposures up-regulate BRCA1 mRNA and reduce mammary tumorigenesis. Carcinogenesis 25(5):741–748
80. Caëtano B, Le Corre L, Chalabi N, Delort L, Bignon Y, Bernard-Gallon DJ (2006) Soya phytonutrients act on a panel of genes implicated with BRCA1 and BRCA2 oncosuppressors in human breast cell lines. Br J Nutr 95 (2):406
81. Hilakivi-Clarke L, Cho E, Clarke R (1998) Maternal genistein exposure mimics the effects of estrogen on mammary gland development in female mouse offspring. Oncol Rep 5(3):609–616

82. Hilakivi-Clarke L, Cho E, Cabanes A, DeAssis S, Olivo S, Helferich W, Lippman MC, Clarke R (2002) Dietary modulation of pregnancy estrogen levels and breast cancer risk among female rat offspring. Clin Cancer Res 8 (11):3601–3610
83. Foster WG, Younglai EV, Boutross-Tadross O, Hughes CL, Wade MG (2004) Mammary gland morphology in Sprague-Dawley rats following treatment with an organochlorine mixture in utero and neonatal genistein. Toxicol Sci 77(1):91–100
84. Latendresse JR, Bucci TJ, Olson G, Mellick P, Weis CC, Thorn B, Newbold RR, Delclos KB (2009) Genistein and ethinyl estradiol dietary exposure in multigenerational and chronic studies induce similar proliferative lesions in mammary gland of male Sprague–Dawley rats. Reprod Toxicol 28(3):342–353
85. vom Saal FS, Nagel SC, Coe BL, Angle BM, Taylor JA (2012) The estrogenic endocrine disrupting chemical bisphenol A (BPA) and obesity. Mol Cell Endocrinol 354 (1–2):74–84
86. Voelkel W, Colnot T, Csanady G, Filser J, Dekant W (2002) Metabolism and kinetics of bisphenol A in humans at low doses following oral administration. Chem Res Toxicol 15 (10):1281–1287
87. Durando M, Kass L, Piva J, Sonnenschein C, Soto AM, Luque EH, Muñoz-de-Toro M (2007) Prenatal bisphenol A exposure induces preneoplastic lesions in the mammary gland in Wistar rats. Environ Health Perspect 115 (1):80–86
88. Markey CM, Luque EH, de Toro MM, Sonnenschein C, Soto AM (2001) In utero exposure to bisphenol A alters the development and tissue organization of the mouse mammary gland. Biol Reprod 65(4):1215–1223
89. Muñoz-de-Toro M, Markey CM, Wadia PR, Luque EH, Rubin BS, Sonnenschein C, Soto AM (2005) Perinatal exposure to bisphenol-A alters peripubertal mammary gland development in mice. Endocrinology 146 (9):4138–4147
90. Murray TJ, Maffini MV, Ucci AA, Sonnenschein C, Soto AM (2007) Induction of mammary gland ductal hyperplasias and carcinoma in situ following fetal bisphenol A exposure. Reprod Toxicol 23(3):383–390
91. Vandenberg LN, Maffini MV, Schaeberle CM, Ucci AA, Sonnenschein C, Rubin BS, Soto AM (2008) Perinatal exposure to the xenoestrogen bisphenol-A induces mammary intraductal hyperplasias in adult CD-1 mice. Reprod Toxicol 26(3):210–219
92. Shelby MD (2008) NTP-CERHR monograph on the potential human reproductive and developmental effects of bisphenol A. NTP CERHR MON (22):v-v
93. Acevedo N, Davis B, Schaeberle CM, Sonnenschein C, Soto AM (2013) Perinatally administered bisphenol a as a potential mammary gland carcinogen in rats. Environ Health Perspect 121(9):1040–1046
94. Mandrup K, Boberg J, Isling LK, Christiansen S, Hass U (2016) Low-dose effects of bisphenol A on mammary gland development in rats. Andrology 4(4):673–683
95. NTP National Toxicology Program (2008) NTP-CERHR monograph on the potential human reproductive and developmental effects of bisphenol A. RTP, NC 27709
96. FDA (2016) Bisphenol A (BPA): use in food contact application. http://www.fda.gov/newsevents/publichealthfocus/ucm064437.htm
97. EFSA. European Food Safety Authority (2015) Scientific opinion on bisphenol A (2015). https://www.efsa.europa.eu/sites/default/files/corporate_publications/files/factsheetbpa150121.pdf
98. Vandenberg LN, Prins GS (2016) Clarity in the face of confusion: new studies tip the scales on bisphenol A(BPA). Andrology 4:561–564
99. Bhargava HN, Leonard PA (1996) Triclosan: applications and safety. Am J Infect Control 24:209–218
100. Halden RU, Paull DH (2005) Co-occurrence of triclocarban and triclosan in U.S. water resources. Environ Sci Technol 39 (6):1420–1426
101. Sandborgh-Englund G, Adolfsson-Erici M, Odham G, Ekstrand J (2006) Pharmacokinetics of triclosan following oral ingestion in humans. J Toxicol Environ Health A 69:1861–1873
102. Adolfsson-Erici M, Pettersson M, Parkkonen J, Sturve J (2002) Triclosan, a commonly used bactericide found in human milk and in the aquatic environment in Sweden. Chemosphere 46:1485–1489
103. Calafat XY, Wong L, Reidy JA, Needham LL (2008) Urinary concentrations of triclosan in the U.S. population: 2003–2004 Antonia M. Environ Health Perspect 116(3):303–307
104. Allmyr M, Adolfsson-Erici M, McLachlan MS, Sandborgh Englund G (2006) Triclosan in plasma and milk from Swedish nursing mothers and their exposure via personal care products. Sci Total Environ 372:87–93

105. Crofton KM, Paul KB, Devito MJ, Hedge JM (2007) Short-term in vivo exposure to the water contaminant triclosan: evidence for disruption of thyroxine. Environ Toxicol Pharmacol 24(2):194–197
106. Gee RH, Charles A, Taylor N, Darbre PD (2008) Oestrogenic and androgenic activity of triclosan in breast cancer cells. J Appl Toxicol 28:78–91
107. Stoker TE, Gibson EK, Zorrilla LM (2010) Triclosan exposure modulates estrogen-dependent responses in the female Wistar rat. Toxicol Sci 117(1):45–53
108. Darbre PD, Charles AK (2010) Environmental oestrogens and breast cancer: evidence for combined involvement of dietary, household and cosmetic xenoestrogens. Anticancer Res 30(3):815
109. Dinwiddie MT, Terry PD, Chen J (2014) Recent evidence regarding triclosan and cancer risk. Int J Environ Res Public Health 11 (2):2209–2209
110. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm378542.htm
111. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm517478.htm
112. US EPA (2000) Reregistration eligibility decision (RED) vinclozolin. Office of Prevention, Pesticides and Toxic Substances. US Environmental Protection Agency
113. Kelce WR, Monosson E, Gamcsik MP, Laws SC, Gray LEJ (1994) Environmental hormone disruptors: evidence that vinclozolin developmental toxicity is mediated by antiandrogenic metabolites. Toxicol Appl Pharmacol 126:276–285
114. Monosson E, Kelce WR, Lambright C, Ostby J, Gray LE Jr (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 15:65–79
115. Gray LE Jr, Ostby J, Furr J, Wolf CJ, Lambright C, Parks L, Veeramachaneni DN, Wilson V, Price M, Hotchkiss A, Orlando E, Guillette L (2001) Effects of environmental antiandrogens on reproductive development in experimental animals. Hum Reprod 7:248–264
116. Christiansen S, Scholze M, Dalgaard M, Vinggaard AM, Axelstad M, Kortenkamp A, Hass U (2009) Synergistic disruption of external male sex organ development by a mixture of four antiandrogens. Environ Health Perspect 117:1839–1846
117. Yeh S, Hu YC, Wang P, Xie C, Xu Q, Tsai M, Dong Z, Wang R, Lee T, Chang C (2003) Abnormal mammary gland development and growth retardation in female mice and MCF7 breast cancer cells lacking androgen receptor. J Exp Med 198:1899–1908
118. Irigaray P, Newby JA, Clapp R, Hardell L, Howard V, Montagnier L, Epstein S, Belpomme D (2007) Lifestyle-related factors and environmental agents causing cancer: an overview. Biomed Pharmacother 61:640–658
119. Guerrero-Bosagna C, Settles M, Lucker BJ, Skinner MK (2010) Epigenetic transgenerational actions of vinclozolin on promoter regions of the sperm epigenome. PLoS One 5:e13100
120. Anway MD, Cupp AS, Uzumcu M, Skinner MK (2005) Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science 308:1466–1469
121. Anway MD, Leathers C, Skinner MK (2006) Endocrine disruptor vinclozolin induced epigenetic transgenerational adult-onset disease. Endocrinology 147(12):5515–5523
122. El Sheikh Saad H, Meduri G, Phrakonkham P, Bergès R, Vacher S, Djallali M, Auger J, Canivenc-Lavier MC, Perrot-Applanat M (2011) Abnormal peripubertal development of the rat mammary gland following exposure in utero and during lactation to a mixture of genistein and the food contaminant vinclozolin. Reprod Toxicol 32(1):15–25
123. El Sheikh Saad H, Toullec A, Vacher S, Pocard M, Bieche I, Perrot-Applanat M (2013) In utero and lactational exposure to vinclozolin and genistein induces genomic changes in the rat mammary gland. J Endocrinol 216(2):245–263
124. Knower KC, To, S. Q, Leung YK, Ho SM, Clyne CD (2014) Endocrine disruption of the epigenome: a breast cancer link. Endocr Relat Cancer 21:T33–T55
125. Rudel RA, Ackerman JM, Attfield KR, Brody JG (2014) New exposure biomarkers as tools for breast cancer epidemiology, biomonitoring, and prevention: a systematic approach based on animal evidence. National Institute of Environmental Health Sciences, United States
126. Emond C, DeVito M, Warner M, Eskenazi B, Mocarelli P, Birnbaum LS (2016) An assessment of dioxin exposure across gestation and lactation using a PBPK model and new data from seveso. Environ Int 92–93:23–32

127. Bruner-Tran KL, Gnecco J, Ding T, Glore DR, Pensabene V, Osteen KG (2016) Exposure to the environmental endocrine disruptor TCDD and human reproductive dysfunction: translating lessons from murine models. Reprod Toxicol 68:59–71
128. Safe SH (1995) Modulation of gene expression and endocrine response pathways by 2,3,7,8-tetrachlorodibenzo-p-dioxin and related compounds. Pharmacol Ther 67 (2):247–281
129. Rogers JM, Denison MS (2002) Analysis of the antiestrogenic activity of 2,3,7,8-tetrachlorodibenzo-p-dioxin in human ovarian carcinoma BG-1 cells. Mol Pharmacol 61 (6):1393–1403
130. Brown N, Manzolillo P, Zhang J, Wang J, Lamartiniere C (1998) Prenatal TCDD and predisposition to mammary cancer in the rat. Carcinogenesis 19(9):1623–1629
131. Fenton S, Hamm J, Birnbaum L, Youngblood G (2002) Persistent abnormalities in the rat mammary gland following gestational and lactational exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). Toxicol Sci 67 (1):63–74
132. Vorderstrasse BA, Fenton SE, Bohn AA, Cundiff JA, Lawrence B (2004) A novel effect of dioxin: exposure during pregnancy severely impairs mammary gland differentiation. Toxicol Sci 78(2):248–257
133. Lew BJ, Collins LL, O'Reilly MA, Lawrence BP (2009) Activation of the aryl hydrocarbon receptor during different critical windows in pregnancy alters mammary epithelial cell proliferation and differentiation. Toxicol Sci 111 (1):151–162
134. Warner M, Mocarelli P, Samuels S, Needham L, Brambilla P, Eskenazi B (2011) Dioxin exposure and cancer risk in the seveso women's health study. Environ Health Perspect 119(12):1700–1705
135. Warner M, Eskenazi B, Mocarelli P, Gerthoux PM, Samuels S, Needham L, Patterson D, Brambilla P (2002) Serum dioxin concentrations and breast cancer risk in the seveso women's health study. Environ Health Perspect 110(7):625–628
136. Leijs MM, Koppe JG, Olie K, Aalderen WMC v, Voogt P d, Vulsma T, Westra M, ten Tusscher GW (2008) Delayed initiation of breast development in girls with higher prenatal dioxin exposure; a longitudinal cohort study. Chemosphere 73(6):999–1004
137. Pup LD, Mantovani A, Cavaliere C, Facchini G, Luce A, Sperlongano P, Caraglia M, Berretta M (2016) Carcinogenetic mechanisms of endocrine disruptors in female cancers (review). Oncol Rep 36(2):603–612
138. IARC (1993) Monographs on the Evaluation of the carcinogenic risk of chemicals to humans. Beryllium, cadmium, mercury, and exposures in the glass manufacturing industry. International Agency for Research on Cancer, Lyon, France
139. Garcia-Morales P, Saceda M, Kenney N, Kim N, Salomon DS, Gottardis MM, Solomon HB, Sholler PF, Jordan VC, Martin MB (1994) Effect of cadmium on estrogen receptor levels and estrogen-induced responses in human breast cancer cells. J Biol Chem 269 (24):16896
140. Martin MB, Reiter R, Trock B, Paik S, Lirio AA, Kenney N, Stoica A, Foss C, Chepko G, Singh B, Hilakivi-Clarke L, Clarke R, Sholler PF, Johnson MD (2003) Cadmium mimics the in vivo effects of estrogen in the uterus and mammary gland. Nat Med 9 (8):1081–1084
141. Davis J, Khan G, Martin MB, Hilakivi-Clarke L (2013) Effects of maternal dietary exposure to cadmium during pregnancy on mammary cancer risk among female offspring. J Carcinog 12(1):11–11
142. Öhrvik H, Yoshioka M, Oskarsson A, Tallkvist J, Sveriges lantbruksuniversitet (2006) Cadmium-induced disturbances in lactating mammary glands of mice. Toxicol Lett 164 (3):207–213
143. Fenga C (2016) Occupational exposure and risk of breast cancer. Biomed Rep 4 (3):282–292
144. Lin J, Zhang F, Lei Y (2016) Dietary intake and urinary level of cadmium and breast cancer risk: a meta-analysis. Cancer Epidemiol 42:101–107
145. Filgo AJ, Quist EM, Hoenerhoff MJ, Brix AE, Kissling GE, Fenton SE (2015) Perfluorooctanoic acid (PFOA)–induced liver lesions in two strains of mice following developmental exposures: PPARα is not required. Toxicol Pathol 43(4):558–568
146. Olsen GW, Burris JM, Ehresman DJ, Froehlich JW, Seacat AM, Butenhoff JL, Zobel LR (2007) Half-life of serum elimination of perfluorooctanesulfonate, perfluorohexanesulfonate, and perfluorooctanoate in retired fluorochemical production workers. Environ Health Perspect 115(9):1298–1305
147. Lau C (2012) Perfluorinated compounds. Springer Basel, Basel, pp 47–86
148. Yang C, Tan YS, Harkema JR, Haslam SZ (2009) Differential effects of peripubertal exposure to perfluorooctanoic acid on

mammary gland development in C57Bl/6 and Balb/c mouse strains. Reprod Toxicol 27(3):299–306

149. Macon MB, Villanueva LR, Tatum-Gibbs K, Zehr RD, Strynar MJ, Stanko JP, White SS, Helfant L, Fenton SE (2011) Prenatal perfluorooctanoic acid exposure in CD-1 mice: low-dose developmental effects and internal dosimetry. Toxicol Sci 122(1):134–145
150. White SS, Stanko JP, Kato K, Calafat AM, Hines EP, Fenton SE (2011) Gestational and chronic low-dose PFOA exposures and mammary gland growth and differentiation in three generations of CD-1 mice. Environ Health Perspect 119(8):1070–1076
151. Tucker DK, Macon MB, Strynar MJ, Dagnino S, Andersen E, Fenton SE (2015) The mammary gland is a sensitive pubertal target in CD-1 and C57Bl/6 mice following perinatal perfluorooctanoic acid (PFOA) exposure. Reprodu Toxicol 54:26–36
152. Zhao Y, Tan YS, Haslam SZ, Yang C (2010) Perfluorooctanoic acid effects on steroid hormone and growth factor levels mediate stimulation of peripubertal mammary gland development in C57BL/6 mice. Toxicol Sci 115(1):214–224
153. Bonefeld-Jorgensen EC, Long M, Bossi R, Ayotte P, Asmund G, Krüger T, Ghisari M, Mulvad G, Kern P, Nzulumiki P, Dewailly E (2011) Perfluorinated compounds are related to breast cancer risk in greenlandic inuit: a case control study. Environ Health 10(1):88
154. US EPA (2016) Drinking water health advisory for perfluorooctanoic acid (PFOA). Office of Water. US Environmental Protection Agency
155. IARC (2016) Monograph on the evaluation of carcinogenic risk to humans: some chemicals used as solvents and in polymer manufacture. International Agency for Research on Cancer, Lyon, France
156. IARC (1999) Monograph on the evaluation of carcinogenic risk to humans: occupational exposure in insecticide application and some Pesticides. International Agency for Research on Cancer, Lyon, France
157. Cooper RL, Laws SC, Das PC, Narotsky MG, Goldman JM, Tyrey EL, Stoker TE (2007) Atrazine and reproductive function: mode and mechanism of action studies. Birth Defects Res B Dev Reprod Toxicol 80 (2):98–112
158. Rayner JL, Wood C, Fenton SE (2004) Exposure parameters necessary for delayed puberty and mammary gland development in Long–Evans rats exposed in utero to atrazine. Toxicol Appl Pharmacol 195(1):23–34
159. Rayner JL, Enoch RR, Fenton SE (2005) Adverse effects of prenatal exposure to atrazine during a critical period of mammary gland growth. Toxicol Sci 87(1):255–266
160. Wetzel L, Luempert LI, Breckenridge C, Tisdel M, Stevens J, Thakur A, Extrom P, Eldridge J (1994) Chronic effects of atrazine on estrus and mammary tumor formation in female Sprague-Dawley and Fischer 344 rats. J Toxicol Environ Health 43(2):169–182
161. Wirbisky SE, Freeman JL (2015) Atrazine exposure and reproductive dysfunction through the hypothalamus-pituitary-gonadal (HPG) axis. Toxics 3(4):414–450
162. US EPA (2003) Memorandum. Review of atrazine cancer epidemiology, DP Barcode D2950200, chemical #080803. Office of Prevention, Pesticides and Toxic Substances. US Environmental Protection Agency
163. Rayner JL, Fenton SE (2011) Atrazine: an environmental endocrine disruptor that alters mammary gland development and tumor susceptibility. Springer New York, New York, NY, pp 167–183
164. Knudson AG (1996) Hereditary cancer: two hits revisited. J Cancer Res Clin Oncol 122 (3):135–140
165. NTP (National Toxicology Program) (2016) Report on carcinogens, 14th edn. U.S. Department of Health and Human Services, Public Health Service, Research Triangle Park, NC. http://ntp.niehs.nih.gov/go/roc14/
166. Nguyen D, Oketch-Rabah H, Illa-Bochaca I, Geyer F, Reis-Filho J, Mao J, Ravani S, Zavadil J, Borowsky A, Jerry D, Dunphy K, Seo J, Haslam S, Medina D, Barcellos-Hoff M (2011) Radiation acts on the microenvironment to affect breast carcinogenesis by distinct mechanisms that decrease cancer latency and affect tumor type. Cancer Cell 19(5):640–651
167. Osborne G, Rudel R, Schwarzman M (2015) Evaluating chemical effects on mammary gland development: a critical need in disease prevention. Reprod Toxicol 54:148–155
168. Vandenberg LN, Schaeberleb CM, Rubinb BS, Sonnenscheinb C, Soto AM (2013) The male mammary gland: a target for the xenoestrogen bisphenol A. Reprod Toxicol 37:15–23
169. Diamanti-Kandarakis E, Jean-Pierre Bourguignon J-E, Gore AC (2009) Endocrine-disrupting chemicals: an endocrine society

scientific statement. Endocr Rev 30 (4):293–342

170. Pinter A, Torok G, Borzsonyi M, Surjan A, Calk M, Kelecsenvi Z, Kocsis Z (1990) Long-term carcinogenicity bioassay of the herbicide atrazine in F344 rats. Neoplasma 37 (5):533–544
171. Brown NM, Lamartiniere CA (1995) Xenoestrogens alter mammary gland differentiation and cell proliferation in the rat. Environ Health Perspect 103:708–713
172. Jenkins S, Rowell C, Wang J, Lamartiniere CA (2007) Prenatal TCDD exposure predisposes for mammary cancer in rats. Reprod Toxicol 23:391–396
173. Radisky DC, Hartmann LC (2009) Mammary involution and breast cancer risk: transgenic models and clinical studies. J Mammary Gland Biol Neoplasia 14:181–191
174. Simpkins JW, Swenberg JA, Weiss N, Brusick D, Eldridge JC, Stevens JT, Handa RJ, Hovey RC, Plant TM, Pastoor TP, Breckenridge CB (2011) Atrazine and breast cancer: a framework assessment of the toxicological and epidemiological evidence. Toxicol Sci 123 (2):441–459
175. IARC (International Agency for Research on Cancer) (1997) Polychlorinated dibenzo-para-dioxins and polychlorinated dibenzofurans. IARC Monogr Eval Carcinog Risks Hum 69
176. Steenland K, Bertazzi P, Baccarelli A, Kogevinas M (2004) Dioxin revisited: developments since the 1997 IARC classification of dioxin as a human carcinogen. Environ Health Perspect 112(13):1265–1268
177. Boffetta P, Mundt KA, Adami HO, Cole P, Mandel JS (2011) TCDD and cancer: a critical review of epidemiologic studies. Crit Rev Toxicol 41(7):622–636. (review)
178. Herbst AL, Ulfelder H, Poskanzer DC (1971) Adenocarcinoma of the vagina: Association of Maternal Stilbestrol Therapy with tumor appearance in young women. N Engl J Med 284(16):878–881
179. Bromer JG, Zhou Y, Taylor MB, Doherty L, Taylor HS (2010) Bisphenol-A exposure in utero leads to epigenetic alterations in the developmental programming of uterine estrogen response. FASEB J 24(7):2273–2280
180. Bromer JG, Wu J, Zhou Y, Taylor HS (2009) Hypermethylation of homeobox A10 by in utero diethylstilbestrol exposure: an epigenetic mechanism for altered developmental programming. Endocrinology 150 (7):3376–3382
181. Haynes BA, Mookadam F (2009) Male gynecomastia. Mayo Clin Proc 84 (8):672–672
182. Felner EL, White PC (2000) Prepubertal gynecomastia: indirect exposure to estrogen cream. Pediatrics 105(4):e55–E55
183. Henley DV, Lipson N, Korach KS, Bloch CA (2007) Prepubertal gynecomastia linked to lavender and tea tree oils. N Engl J Med 356 (5):479–485
184. Thayer KA, Foster PM (2007) Workgroup Report: National Toxicology Program Workshop on hormonally induced reproductive tumors: relevance of rodent bioassays. Environ Health Perspect 115(9):1351–1356
185. Rudel RA, Attfield KR, Schifano JN, Brody JG (2007) Chemicals causing mammary gland tumors in animals signal new directions for epidemiology, chemicals testing, and risk assessment for breast cancer prevention. Cancer 109:2635–2666
186. Cohn BA, Wolff MS, Cirillo PM, Sholtz RI (2007) DDT and breast cancer in young women: new data on the significance of age at exposure. Environ Health Perspect 115 (10):1406–1414
187. Ly D, Forman D, Ferlay J, Brinton LA, Cook MB (2013) An international comparison of male and female breast cancer incidence rates. Int J Cancer 132(8):1918–1926
188. Skinner MK (2007) Endocrine disruptors and epigenetic transgenerational disease etiology. Pediatr Res 61:48R–50R

Methods in Pharmacology and Toxicology (2017) 235–259
DOI 10.1007/7653_2016_67
© Springer Science+Business Media New York 2016
Published online: 24 December 2016

Alternative Methods Used to Assess Potential Embryo-Fetal Developmental Risk of Pharmaceuticals

Maia L. Green, Joshua T. Gamse, and Kimberly C. Brannen

Abstract

Alternative developmental toxicity assays are used by pharmaceutical companies to detect the teratogenic potential of human drugs. These methods are intended to reduce, refine, or replace (3Rs) animal use in nonclinical embryo-fetal developmental toxicity testing. Screening methods (e.g., rodent whole embryo culture, embryonic stem cells, and zebrafish) are powerful tools to identify hazards; they also provide unique mechanistic insights that improve our understanding of developmental toxicology. By improving the sophistication of these models over the past two decades, the field of developmental and reproductive toxicology has been preparing to meet the increased need for enhanced developmental toxicity testing in drug discovery and development. Interest in these tools has been further amplified since the International Conference on Harmonisation (ICH) indicated that it will address the use of in vitro, ex vivo, and non-mammalian developmental toxicity assays for regulatory purposes in the upcoming revisions to the ICH S5 (R2) guideline. Moreover, alternative assays combined with newer technologies such as high-content imaging, automated embryo handling, and functional genomics may expedite testing of pharmaceuticals for teratogenic liabilities while also increasing the informational content of these screens. To illustrate the potential of alternative developmental toxicity assays, some examples of modern methods as well as modifications to enhance these methods in the future are highlighted.

Keywords: Alternative assays, Developmental toxicity, Embryo, Teratogen

1 Introduction

Nonclinical embryo-fetal developmental (EFD) toxicity studies are used to identify potential risks to the developing embryo/fetus in pregnant women that may receive a new pharmaceutical. Because little or no human developmental toxicity data are available when a new drug enters the market, these nonclinical data in animal models are particularly important. A teratogenic liability is highly undesirable and can be unacceptable in general medicine programs. A quick, low-cost screen for teratogenicity could positively impact clinical trial enrollment, improve product labeling, and reduce business costs. However, predicting developmental toxicity is a complex and difficult task. Developing conceptuses (and, to some extent, the maternal environment in which they grow) represent many coordinated systems and pathways that function together and

change considerably over time which makes this a complex system to model. Therefore, in order to enhance the capability of drug development, there is a need for predictive screening methods to be fast, simple, and inexpensive with enough biological complexity to capture all or most of the potential vulnerabilities to developmental toxicity.

A number of methods were developed and evaluated for this purpose in the late twentieth and early twenty-first centuries. These included tissue culture (e.g., embryonic stem cells, micromass), embryonic limb bud organ culture, whole embryo culture (rodent or rabbit), and non-mammalian models (e.g., hydra, frog embryo teratogenesis assay xenopus [FETAX], zebrafish). Of these models, the three that are most commonly used at the time of writing this (2016) are whole embryo culture, embryonic stem cells, and zebrafish, each of which is described in more detail in the sections below. Assays with these three models have been found to have fairly robust predictivity for developmental toxicity [1–11], which can make them useful for hazard identification and internal decision-making by pharmaceutical companies. However, it is important to consider how predictivity for these models is defined. Traditionally, two embryo-fetal developmental (EFD) toxicity studies are conducted (one study in a rodent and the other in a non-rodent mammalian species) [12], and these species serve as models to predict human toxicity. Alternative models must be evaluated based on their ability to predict in vivo mammalian outcomes, but because developmental toxicity outcomes are often not the same among all species (e.g., rat, rabbit, human), investigators must determine which species they aim to predict. While it would be ideal to predict human outcomes, evaluating a model by this standard is very difficult given the scarcity of human developmental toxicity data. Therefore, many researchers choose to use the outcomes in one or more traditional mammalian in vivo models as the "true" results against which to assess predictions from alternative models. Then the performance of alternative models can be evaluated by comparing data for pharmaceuticals in the traditional model(s) to results from the alternative assay. It may even be possible to discover patterns of concordance and discordance to understand the source of discrepancies between outcomes in alternative and traditional assays.

While these methods are unlikely to replace traditional in vivo mammalian studies altogether, it is becoming increasingly clear that they have gained substantial traction in the pharmaceutical industry and that the impact on drug development may extend beyond lead optimization. For example, the International Conference on Harmonisation (ICH) has signaled its intent to address the possible use of in vitro, ex vivo, and non-mammalian developmental toxicity assays for regulatory purposes in the upcoming revisions to the ICH S5(R2) guideline: detection of toxicity to reproduction of

medicinal products and toxicity to male fertility. How much and under what circumstances these alternatives will affect regulatory decisions still has to be determined, but it seems apparent that for the foreseeable future, alternative assays will be an integral part of pharmaceutical developmental toxicity testing and decision-making for many sponsors.

A vision for the future of developmental toxicity testing would be alternative model systems in which a large array of pharmaceuticals can be rapidly tested on large numbers of developing embryos, followed by high-content analysis of morphology that precisely evaluates the effects of the pharmaceuticals, and finally tools that allow exposure-based predictions [13] and mechanism(s) of teratogenicity to be elucidated. New technologies in automated embryo handling, high-content imaging, and functional genomics are improving the speed of screening pharmaceuticals for teratogenic liabilities while also increasing the informational content of these screens. The progress toward this vision and the limitations that remain will be reviewed in the context of developmental toxicity testing.

2 Alternative Mammalian Models

2.1 Whole Embryo Culture (WEC)

The rodent whole embryo culture (WEC) assay is an in vitro screening tool that was designed in the 1970s by New [14], and the methodology has been modified since its original development [15–17]. WEC can predict in vivo potential teratogenicity for pharmaceuticals and other test substances [18, 19] (Fig. 1). The rat WEC assay performance for predicting in vivo teratogenicity has been demonstrated in studies comparing in vitro WEC results with in vivo EFD toxicity outcomes [5, 8, 20]. This is notable because the rat is generally the primary test species used to evaluate the potential developmental and reproductive toxicity of drug candidates, making this a compelling model.

Within this in vitro system, postimplantation mouse or rat conceptuses are removed from the uterus (3–5 somites) [21]. The decidua and Reichert's membrane are removed, but the visceral yolk sac remains intact. Embryos are placed in culture bottles that contain a high percentage of serum (up to 100%) to which the pharmaceutical is added. The embryos are maintained in a rotating culture unit with the stage-appropriate concentrations of oxygen at 37 °C for up to 72 h. The developmental stages are highly sensitive to teratogen insult because morphogenic processes controlling organ system development are occurring. The embryo continues to grow within the surrounding amniotic fluid and visceral yolk sac during the culture period. The culture consists of a humidified gas atmosphere composed of progressively increasing levels of oxygen (5% until 5 h post-culture (hpc); 10%, 5–21 hpc; 20%, 21–45 hpc;

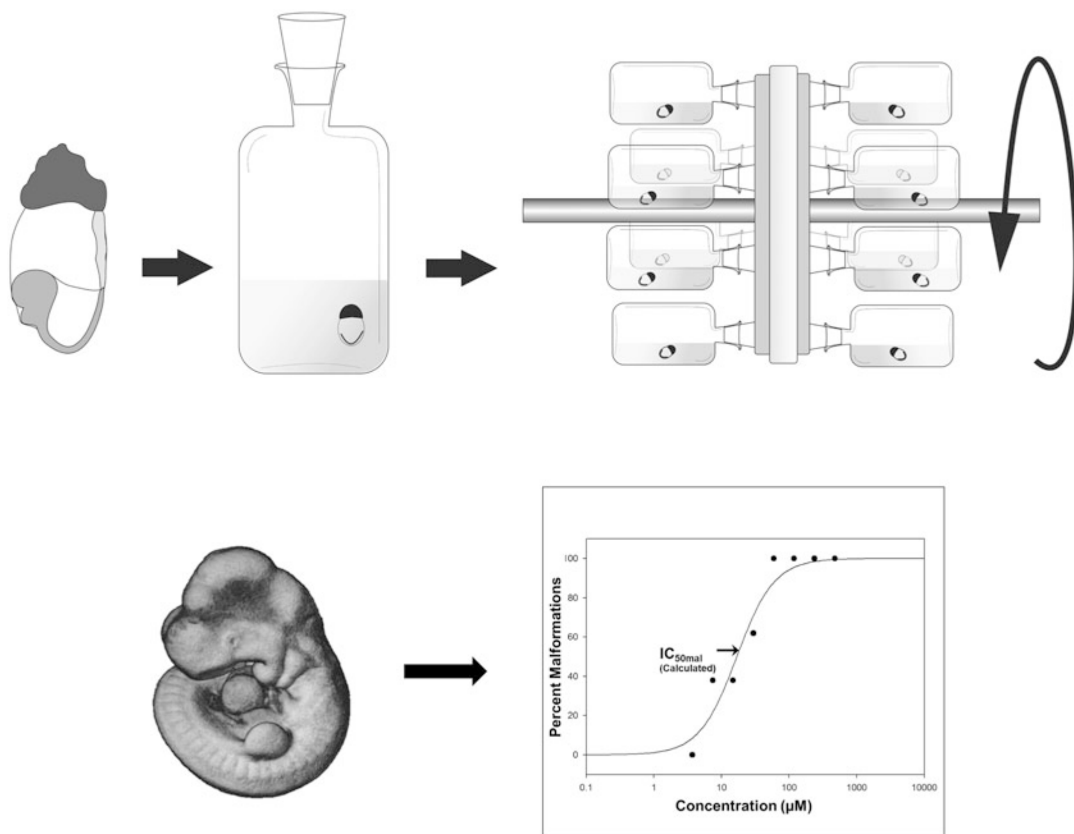

Fig. 1 Rat whole embryo culture assay evaluates drug effects after a 48-h incubation period during mid-stage embryo development. Embryos are typically removed from the uterus on gestation day 9 (early somite stage), and the decidua and Reichert's membrane are dissected away. The visceral yolk sac and ectoplacental cone are left intact. Then embryos are placed in bottles with medium (primarily rat serum containing various concentrations of the pharmaceutical being tested) that is gassed with 5% oxygen. These bottles are placed on a rotator and maintained at 37 °C in an incubator for up to 48 h, with increasing oxygen concentrations appropriate for the developmental stage. At the end of the culture period, embryo morphology is evaluated for control and treated groups, and endpoints such as an IC_{50mal} (concentration at which 50% of embryos are malformed) are calculated (reprinted from Brannen et al. [61] with permission from Oxford University Press)

and 40%, 45+ hpc), 5% carbon dioxide, and a nitrogen balance for the duration of the culture period to compensate for embryo growth. At the termination of the culture period, embryos are removed from the serum and routinely assessed for growth, viability, and gross anatomical abnormalities via examination with a microscope. Each embryo is assessed and scored based on developmental stage. A number of embryonic scoring systems (e.g., [21–25]) have been developed to quantitate the morphologic changes that may represent precursors to in vivo teratogenicity.

To identify molecular precursors, transcriptional profiles in WEC can be used to highlight biological mechanisms underlying phenotypic changes. For instance, the gene expression response

after valproic acid exposure in WEC was compared with responses in other in vitro developmental toxicity models. The results were complimentary in that similar biological processes were elevated from the dataset [26]. This emphasized the use of a battery of assays to optimize the detection of developmental toxicants.

The rabbit is often selected as a second test species for in vivo developmental toxicity testing. A rabbit WEC assay was designed by Naya and colleagues, who reported successful in vitro development of gestation day 9 and gestation day 10 rabbit embryos [27]. The methodology was later improved, and a modified scoring system for the WEC was published [28] to align with the Brown and Fabro scoring system.

The utility of using both rodent and rabbit WEC assays for predicting in vivo developmental toxicity was tested in a study comparing results from thalidomide-treated WEC with thalidomide EFD outcomes [28]. Thalidomide induces limb reduction defects in fetal rabbits, whereas rats are insensitive. When thalidomide was added to the culture media that contained rat or rabbit conceptuses, neither in vitro assay revealed changes in morphology, growth, or viability parameters [29]. However, using the WEC system, the authors identified glutathione (GSH), an antioxidant, as a potential inherent biochemical difference between the two species because levels were 50% lower than controls in thalidomide-treated rabbit embryos compared to no change in thalidomide-treated rat embryos. This suggested that the differential susceptibility of thalidomide teratogenesis was at least partially due to low GSH levels because of the high levels of DNA oxidation.

2.2 Advances in Whole Embryo Imaging Using Micro-Computed Tomography (Micro-CT)

The microscopic examination process of the WEC assay is low throughput. However, there is potential that this process could be enhanced or replaced by utilizing imaging techniques, such as micro-CT. Advantages of scanning embryos to obtain electronic images instead of manual morphological assessments by microscope include easier viewing of digital specimens, reduced operator time, and increased throughput. In addition, virtual histology (high-resolution electronic images of embryonic sections used for morphological assessments) allows the investigator to make detailed assessments without additional histological methods (Fig. 2 [30]). In addition, the embryo images can be stored electronically and assessed later.

Micro-CT, a 3-D imaging modality based on the attenuation of X-rays, is a common choice for 3-D imaging of soft tissue when combined with contrast staining. Compared to MRI and optical imaging, it is particularly suitable for small samples such as embryos for high resolution and high throughput [31–33]. This platform is applicable to both embryo and fetal morphological assessments for in vitro WEC and in vivo EFD studies, respectively [34]. For instance, micro-CT was successfully used for quantitative image

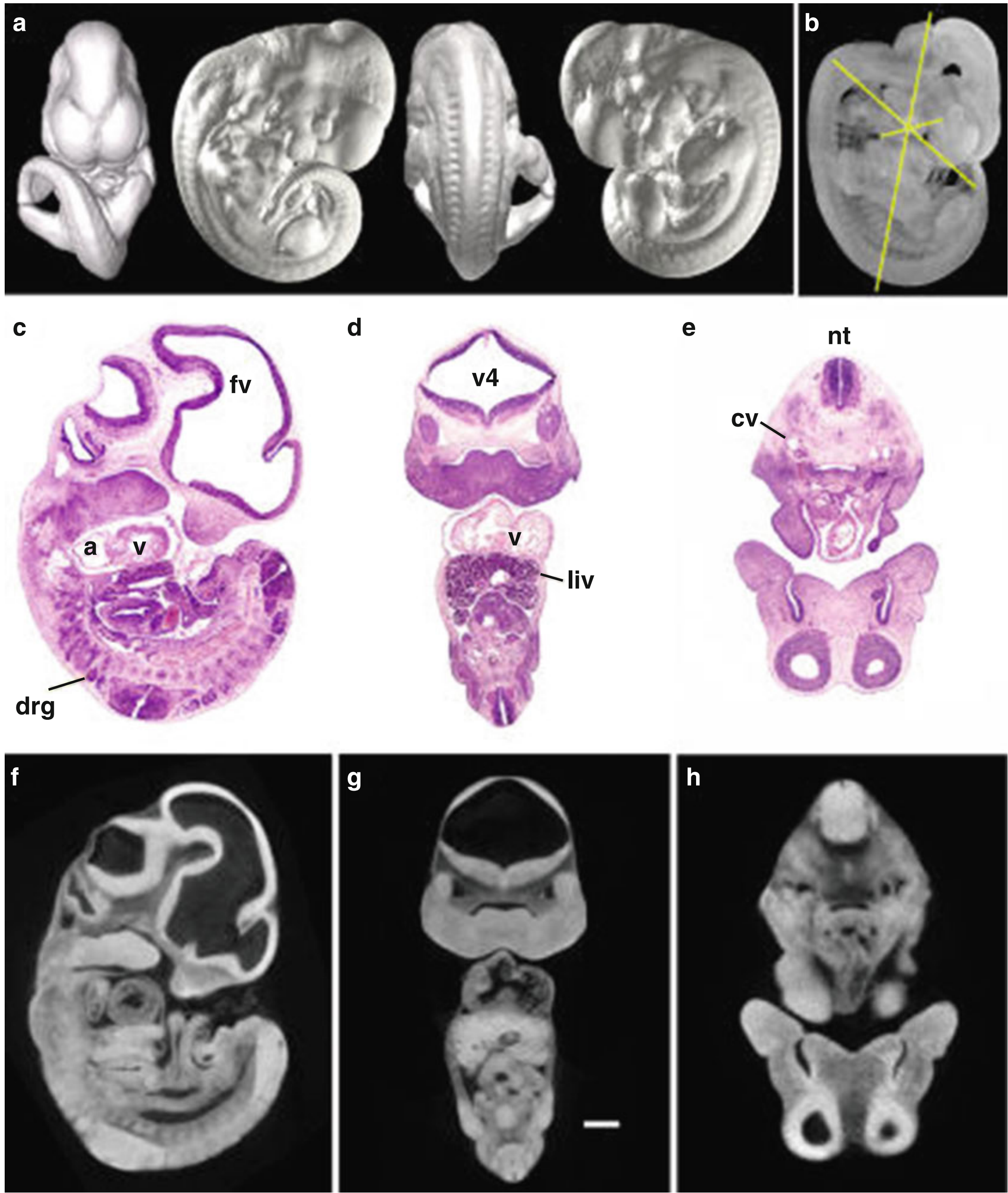

Fig. 2 Comparison of paraffin and virtual histology of micro-CT E11.5 mouse embryos. Embryos were scanned at 8 μm with isosurface renderings in the *top row* (**a**, **b**), traditional histology sections from a littermate in the *middle row* (**c**–**e**), and the virtual histology via micro-CT in the *bottom row* (**f**–**h**). *a* cardiac atrium, *cv* cardinal vein, *drg* dorsal root ganglia, *fv* forebrain vesicle, *liv* liver, *nt* neural tube, *v* cardiac ventricle, *v4* fourth ventricle (modified from Johnson et al. [30])

analysis of live avian embryos [35]. A useful review of micro-CT imaging methods for embryos was published [36]. One method uses a hydrogel scaffold to mechanically support the embryo for imaging. The resulting embryonic tissues are permeable to iodine staining and optically transparent [37]. However, tissue shrinkage and deformation may occur during the hypertonic iodine staining [38]. To overcome this, a hydrogel stabilization approach to support tissue structure and preserve conformation was developed [39]. As a result, high-quality embryo images with minimal tissue distortion can be achieved along with virtual histology. Micro-CT is emerging as a powerful tool for studying embryonic development and quantifying anatomical and physiological changes of the embryo.

2.3 Mouse Embryonic Stem Cell Test (mEST)

The mEST is another type of in vitro assay used for pharmacological screening to detect potential developmental toxicity [9, 40]. The system assesses early embryonic development when embryonic stem cells (ESCs) begin to differentiate along endodermal, mesodermal, and ectodermal lineages. ESCs, placed in culture under certain specified conditions, have the ability to transition from a pluripotent to a more differentiated state, thus recapitulating some in vivo embryonic processes. An mEST was originally designed and validated by the European Centre for the Validation of Alternative Methods and involved culturing the D3 ESC line for 10 days and evaluating the presence of differentiated, beating cardiomyocytes [41–43]. However, since then, a number of ESC differentiation methods have been described [42, 44]. Advantages to the assay include use of molecular- and morphological-based endpoints, low test article use, higher throughput, and the ability to maintain ESCs in vitro for an extended period. Hence, the mEST assay has emerged as a valuable screening tool for hazard identification [2, 45].

ESC lines are established from the inner cell mass of the 3.5-day mouse blastocyst. They are cultured and pharmaceuticals are added to the media for up to 10 days. These assays evaluate pharmaceuticals based on the assessment of toxicological endpoints such as inhibition (cytotoxicity), altered differentiation (e.g., beating cardiomyocytes, differentiated neuronal cells), or transcriptional changes reflecting molecular endpoints (i.e., pluripotency and cell lineage markers) which are expressed during gastrulation.

Transcriptomics may be used to enable a detailed assessment of pharmaceutical-induced changes in the molecular messenger RNA (mRNA) and/or microRNA (miRNA) endpoints that are normally expressed when ESCs differentiate along cell lineages to recapitulate early development in vivo. A RT-PCR-based gene expression analysis can be performed against a panel of target genes whose up- or downregulation is identified to predict potential developmental toxicity in rodents. ESCs with altered temporal expression profiles

of developmental regulator genes could give insight into specific mechanism(s) of teratogenesis for a given pharmaceutical [46, 47].

In addition to molecular profiling, a simpler in vitro system for predicting changes in embryo morphogenesis can be used with various mouse ESC lines [47–49]. With a hanging drop or similar technique, ESCs spontaneously differentiate and form three-dimensional (3-D) aggregates called embryoid bodies (EB). EBs mimic cell differentiation during early mammalian embryogenesis and give rise to more mature cells of the three germ layers. Dysregulation of cellular events during any stage of embryogenesis may lead to morphogenic disturbances that could ultimately result in a birth defect. Hence, this system may serve as a unique tool to screen for developmental toxicity by observing morphological changes after treatment with pharmaceuticals. For instance, when pluripotent P19C5 mouse embryonal carcinoma stem cells that mimic the process of gastrulation and axial body elongation of embryos were treated with developmentally toxic pharmaceuticals at concentrations below cytotoxic levels, the resulting EBs were spherical rather than elongated in shape (measured by EB area) over 4 days of culture (Fig. 3). Analyses of P19C5 cells suggested that the elongation morphogenesis of these EBs represented gastrulation and convergent extension along the anterior-posterior body axis [48, 49] and was not evident when ESCs were treated with embryotoxicants. Data suggest that the EB morphology model may provide an easier system to screen for developmental toxicities caused by pharmaceuticals [50]. Advances in automated imaging platforms could allow investigators to quickly access high-throughput image capture and perform morphological assessments of EBs.

2.4 Advances in High Content Embryoid Body Imaging

Traditional image-based systems have limited throughput capability, and therefore microscope-based, high-content instruments are being used for screening cellular toxicity assays. A number of advanced 3-D imaging platforms are available for researchers to eliminate manual qualitative data collection. For example, a recent report showed that high-throughput toxicity screening in 384-well plates with a quick scan time can enhance data acquisition and analysis time. The combination of image processing methodology and large depth-of-field cytometric scanning provides the potential for the enhanced development of the EB morphology model. However, imaging EBs has an important limitation [51]. Although imaging results provided a high-resolution view of spatiotemporal dynamics of early mesodermal differentiation within the EB, they did not provide information on the EB's differentiation stage or the genes driving and driven by the EB progression from mesoderm expansion to the next stage. This emphasizes the use of a combination of endpoints (e.g., 3-D embryo morphology imaging and molecular profiling) to aid in an enhanced data interpretation.

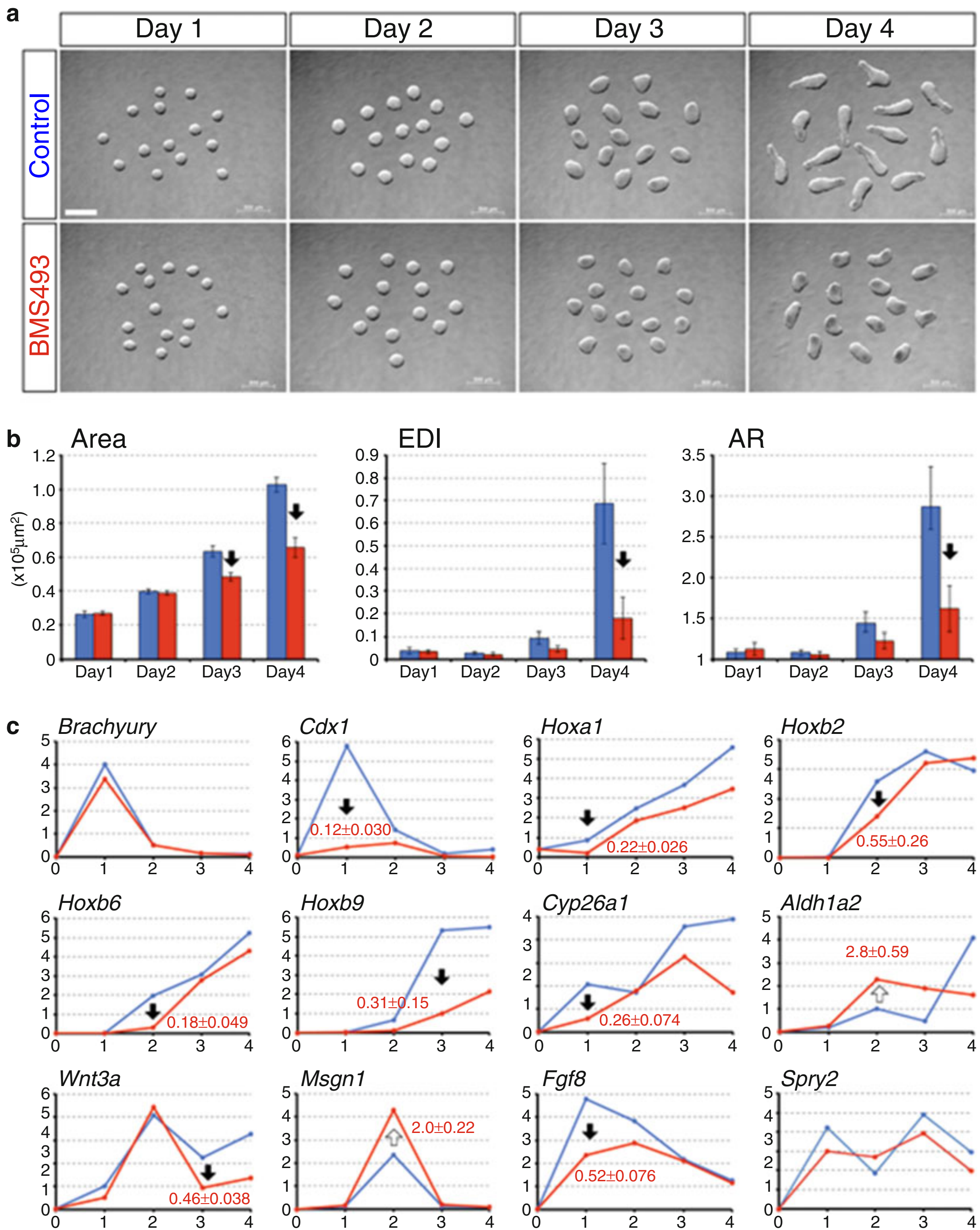

Fig. 3 Impact of retinoic acid signaling inhibition on EB development. EBs were cultured in hanging drops containing 1 μM BMS493, and their morphologies and gene expression profiles were compared to control EBs. (**a**) Images of EBs over 4 days of culture. Scale bars, 500 μm. (**b**) Comparisons of EB morphometric parameters between control EBs (*blue bars*) and BMS493-treated EBs (*red bars*). Each bar represents the mean ± standard deviation ($n = 14$–16) of the morphometric parameters. Graphs show data from one set of experiment. *Downward solid arrows* indicate statistically significant reductions in all three sets of experiments conducted ($P < 0.01$; Student's *t*-test). (**c**) Comparisons of temporal gene expression profiles between control EBs (*blue lines*) and BMS493-treated EBs (*red lines*). Vertical axis represents relative transcript abundance and horizontal axis represents days of culture. Graphs show data from one set of experiments. *Downward solid arrows* and *upward open arrows* indicate consistent reductions and increase, respectively, in all three sets of experiments conducted. Mean ± standard deviation is indicated. (reprinted from Li and Marikawa [50] with permission from John Wiley and Sons)

2.5 Human Embryonic Stem Cell Test (hEST)

The hEST may be utilized to aid in the early discovery-phase detection of potential human developmental toxicants. Metabolomics and transcriptomics can be applied to capture relevant molecular endpoints that are species-specific which may not exist in preclinical species. These molecules could serve as biomarkers of human developmental toxicity at clinically relevant doses. Hence, this test system fills a gap in developmental toxicity screening whereby it can potentially identify human teratogens that animal-based tests do not detect.

There are only a handful of available human ESC (hESC) lines (i.e., H1, H7, H9, H13, and H14) that are well characterized and routinely used in research. These human ESCs are pluripotent and are capable of forming all three germ layers which makes them suitable for human developmental biology research and pharmaceutical screening. An example of a widely used human ESC line is the H9 line, which was first derived from the inner cell mass of blastocysts [52], retained a normal XX karyotype after 6 months of culture, and was passaged continuously for more than 8 months.

The utility of metabolomics in biomarker discovery has been demonstrated with human ESCs [53–56]. Methods utilized plated H9 human ESCs that were allowed to attach and grow for 3 days. Then, the cells were treated with a number of non-teratogenic and teratogenic test agents for approximately 4 days. Media was collected after the treatment period and acetonitrile was added to the samples. Samples were analyzed by LC-MS to measure changes in levels of small molecules in response to test agent exposure. Overall high accuracy was reported with this developmental model [54, 55], although only limited test sets were evaluated. In one of the models, two biomarkers, ornithine and cysteine, were used as indicators of developmental toxicity (77% accuracy with high concordance with in vivo models) when tested using 46 pharmaceuticals, 12 of which were known human teratogens [54]. Nonetheless, this methodology allows potential for predictive modeling and mechanistic understanding of biochemical pathways underlying human developmental toxicity.

Like the mEST assay, expression profiling of human ESCs can be used to identify potential developmental toxicants. Human ESCs can be used as a model of biological systems to assess the impact of pharmaceuticals on the genome expression. The resultant mRNA and microRNA (miRNA) transcriptomes can be assessed by RNA-Seq and TaqMan, respectively. Development of a gene signature (a group of genes in a cell whose combined expression pattern is uniquely characteristic of teratogenicity) using mRNA and/or miRNA endpoints could aid in interpreting the data. A number of potential RNA biomarkers could highlight differences between teratogens. In the future, this could potentially be translated to human concentrations by comparing teratogens at doses at or above the human exposure (AUC or C_{max}) to non-teratogens at

doses below the human exposure. For instance, a gene signature response, initiated at doses below the teratogenic human dose level and exposure, could suggest a promise for predicting a safe clinical therapeutic window.

A major challenge to the implementation and use of the hEST assay as a screening tool for human teratogenicity is its applicability as discussed by Cezar [57]. The assay is difficult to qualify since there are a limited number of known human teratogens. Therefore, conversations surrounding the applicability of the hEST assay in pharmaceutical risk assessment are warranted.

3 Alternative Non-mammalian Models

Several non-mammalian models have been used to study developmental toxicity. The one that is best characterized and is most often used for pharmaceutical research today is the zebrafish. There are many reasons for the zebrafish's popularity in research. Because embryonic and larval development is rapid [58–61], the turn-around time for developmental toxicity screening studies is short (Fig. 4). Zebrafish are also inexpensive, easy to maintain, able to produce large numbers of offspring, transparent, largely genetically homologous to humans, amenable to many research techniques, and useful for mechanistic studies due to the availability of relevant functional genomics and other experimental tools [60, 62–65]. Because zebrafish fertilization and development are external, it is not only feasible to maintain a modest breeding colony of adults for regular production of large numbers of embryos but also easy to grow and observe embryos and larvae under various experimental conditions [62, 63]. This also makes it possible to test the direct effects of test substances on development. Unlike mammalian in vitro and ex vivo models, the zebrafish is a whole-organism, in vivo model with which one can readily assess the entire developmental period and all developing tissues [64, 65]. For these reasons, zebrafish embryos can be used for simple, inexpensive, medium-throughput assays to evaluate most stages and processes of vertebrate development. That has made the model attractive to researchers looking for models to serve as predictive developmental toxicity assays or to study mechanisms of abnormal development, each of which is discussed below.

3.1 Zebrafish Developmental Toxicity Assays

A number of labs have reported the results of their efforts to design and characterize a predictive zebrafish developmental toxicity assay (ZDTA) [3, 6, 7, 10, 11, 66–70]. Several detailed descriptions of assay methodology have also been published [3, 63, 71, 72]. In general, the methods used involve exposing embryos to test substances for a defined window of development, typically starting shortly after fertilization and continuing through the end of the

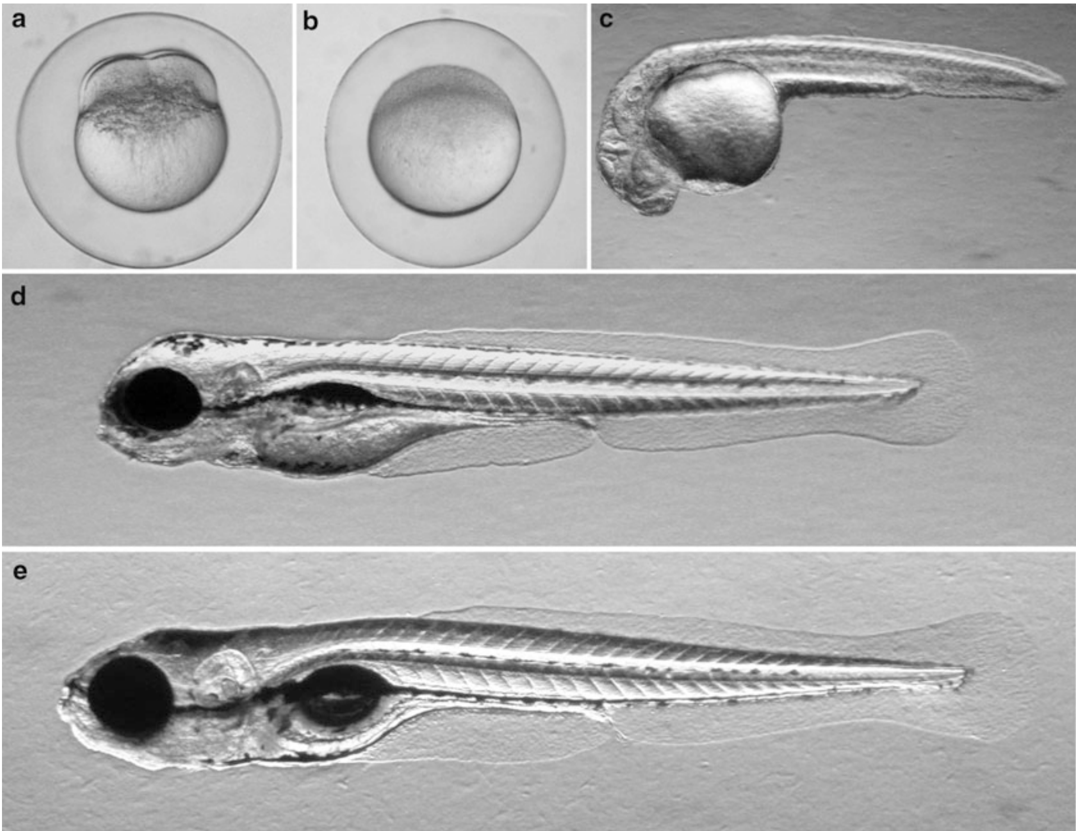

Fig. 4 Zebrafish embryos and larvae at stages included in a typical developmental toxicity assay. (**a**) Embryos are collected from egg traps shortly after spawning (e.g., 1 hpf), rinsed, and selected for use. (**b**) Late-blastula- to early-gastrula-period embryos (approximately 4–6 hpf) are placed into multiwell culture plates filled with buffer (embryo medium) containing several concentrations of the test pharmaceutical. (**c**) A 24-hpf embryo has been removed from the chorion to facilitate viewing and illustrate how rapidly embryonic development occurs; early eye, brain, heart, somite, and notochord development are easily viewed in transparent embryos. (**d**) Hatching is complete and the embryonic period ends at 3 dpf. (**e**) A typical developmental toxicity assay with zebrafish involves evaluation of viability, growth, and morphology of early stage larvae (e.g., 5 dpf) (reprinted from Brannen et al. [61] with permission from Oxford University Press)

embryonic period or early larval period. Embryos are collected within a few hours of fertilization (usually at stages between the cleavage and gastrula periods), cleaned, visually evaluated for suitability (e.g., fertilization, stage, absence of damage), and arrayed among the treatment groups. They are grown in culture medium (a buffer solution) containing the appropriate concentration of the test substance and incubated at approximately 28 or 28.5 °C. After the desired period of exposure, which is usually 3–6 days, larvae are evaluated for viability, growth, and/or morphology. In some cases, embryos/larvae are evaluated at multiple developmental stages during the assay exposure period. Some assays include an evaluation

of developmental toxicity relative to cytotoxicity in an effort to improve the prediction of selective developmental toxicity.

Although there are many similarities in methods used among the various labs doing ZDTAs, there are a few areas that have been somewhat variable. As indicated above, the precise timing of the exposure period and/or data collection can vary. The endpoints included in embryo-larval evaluations and the level of detail with which they are assessed can differ, and this may affect the estimates of the assay's sensitivity and specificity [3, 7]. In addition, zebrafish researchers often disagree about the importance or wisdom of removing the chorion, the acellular envelope that surrounds the embryo for the first few days of development, prior to test substance exposure. Although the chorion has large pores that would be expected to allow small molecules to cross it, concerns about the possibility that the chorion could act as a pharmaceutical sink, slightly delay exposure to the embryo proper, or cause artifacts by interfering with hatching [73–75] have led some to remove it, while others avoid dechorionation due to evidence that it can slightly increase the background rates of mortality and dysmorphology [76].

Given the interlaboratory differences in methodology, it is often difficult to compare results among studies. For example, a number of labs have found concordance between mammalian and zebrafish assays to be over 80% with relatively low false-positive and false-negative rates [3, 7, 11, 69, 77], while the results of other studies have not been as strong [6, 10, 67, 68, 78]. The disparities in concordance among the various studies are likely due to one or more of the following: different assay methods, the selection of pharmaceuticals tested, and different approaches to analyzing the resulting data. It is, therefore, easy to understand why many have advocated for harmonization, and some have attempted to determine an optimal design that could be used in a harmonized manner [6, 10]. However, standard use of a harmonized protocol across labs has remained an elusive goal. Despite the lack of standardization, some pharmaceutical companies have found that zebrafish assays for developmental toxicity and other areas of safety assessment can be valuable for internal decisions (i.e., hazard identification and prioritization of candidate pharmaceuticals) [3, 10, 11, 79].

There are a few limitations to consider regarding the use of zebrafish as an alternative model in pharmaceutical discovery. First, because zebrafish fertilization and development occur externally, there is no maternal component in ZDTAs. While this provides some important logistical benefits as described above, it also means that the role of maternal effects (e.g., metabolism, physiology, pregnancy maintenance) in developmental toxicity cannot be assessed. Of course, the same is also true for mammalian in vitro and ex vivo assays, but this limitation may necessitate the incorporation of data from other sources in order to derive the full benefit of

using predictive screens like ZDTA and others. In addition, there are sometimes concerns about translation of results from a non-mammalian model to humans for risk assessment; while there is a good deal of conservation in genetics and developmental and toxicological processes [80], more needs to be known about how much the differences that exist between zebrafish and humans may affect the ability to predict human outcomes. Finally, the fact that the zebrafish is an aquatic species can be a logistical challenge with pharmaceuticals that have limited aqueous solubility, although it is possible to microinject insoluble pharmaceuticals directly into the yolk cell (see below).

Ultimately, the greatest utility from assays of this type will be realized when—and if—the assay can predict maternal exposure levels that are likely to be safe and those that are likely to carry a risk for developmental toxicity [13]. In order to reach such a lofty goal, it will first be necessary to gain a better understanding of pharmacokinetics and metabolism in the model. Internal exposure levels (tissue concentrations of the test substance) in zebrafish embryos/larvae relative to the concentrations in the culture medium probably vary greatly from one pharmaceutical to another, and it is likely that evaluating internal exposure can improve assay predictivity and translatability [6, 67, 81]. However, few studies include bioanalysis for internal embryo-larval exposure, and those that do usually involve only measurement of whole-larvae samples at a single timepoint. So far, bioanalysis at the earliest stages of ZDTA has been logistically impractical, which limits the questions that can be addressed. There is also much that is still unknown about metabolism in the developing zebrafish, but drug detoxification and activation, especially through maternal metabolism, play important parts in developmental toxicity of some pharmaceuticals. There is at least some metabolic capacity in the liver of embryos and early larvae [82–85], and some proteratogens have been successfully tested in a zebrafish assay [86]. In addition, even when endogenous metabolism is not sufficient to produce a relevant effect, exogenous metabolism approaches may be employed to test the effects in the presence of metabolism [86, 87]. Further work is needed to understand pharmacokinetics and metabolism in zebrafish developmental toxicity assays adequately. When those data are available, it may be possible to make exposure-based predictions from ZDTA results, but at a minimum, the data from zebrafish will at least become even more effective than it already is.

3.2 Neurobehavioral Testing in Zebrafish Larvae

Zebrafish have also become a popular model for studying the processes and pathways involved in neurobehavioral function. Drugs can be screened for developmental neurobehavioral effects in larvae using learning and memory, locomotor, and sensory assays that are analogous to the types of testing performed in pre- and postnatal developmental toxicity and juvenile toxicity studies with

standard mammalian models. Detailed descriptions of zebrafish central nervous system development and function and neurobehavioral testing are beyond the scope of this chapter, but excellent reviews of these topics can be found elsewhere (e.g., see [88, 89]).

3.3 Automation for Rapid Handling of Zebrafish Embryos

As noted earlier, zebrafish have rapid embryonic development, and adults can produce large numbers of synchronously developing embryos [90]. These characteristics make them particularly well suited for high-throughput screening of pharmaceuticals in early stages of discovery: for instance, to narrow down a large number of lead pharmaceuticals by identifying those that present the least hazard to developing embryos and fetuses. Working with so many embryos at once presents an obvious problem, namely, how to precisely and reproducibly manipulate the embryos without damaging them or inhibiting development.

First, large numbers of embryos must be generated and arrayed into microtiter plates. The first issue, spawning of thousands of embryos at the same time, has recently been addressed by commercial zebrafish husbandry companies including Pentair (Mass Embryo Production) and Tecniplast (iSpawn), which use funnel-type chambers to breed tens of fish pairs at one time and quickly collect the embryos. Arraying these embryos into 96-well plates, while avoiding dead embryos, has been automated by various means. Union Biometrica has succeeded, for example, using their COPAS FP large particle flow cytometer. Using optical density measurement, live embryos, which are more transparent than dead ones, can be rapidly arrayed at a desired density into the wells of a microtiter dish. However, the embryos are still in their chorions, which can impede treatment by acting as a sink for certain pharmaceuticals. A robotic embryo loader that incorporates a semi-automated dechorionation step has been developed [74] although it cannot sort live versus dead embryos.

Next, a precise dose of each pharmaceutical to be tested must be administered to each well. This is similar to delivery of chemicals from a library to cultured cells. In many cases this is performed by a robotic liquid handler with micropipette tips that performs a serial dilution prior to adding chemicals to the embryo media. A recent effort to optimize zebrafish embryo dosing in a high-throughput format found that digital dispensing directly into the experimental chambers was significantly better at reproducible treatment [64].

Not every pharmaceutical is amenable to administration in this fashion, as mentioned earlier. Many recently developed pharmaceuticals are highly hydrophobic, making it challenging to treat by diluting in aqueous media, or are large molecules which cannot diffuse into the embryo from the media. An alternative method for treating zebrafish embryos that avoids these constraints is to micro-inject the pharmaceutical directly into the embryo. An automated system that uses high-speed immobilization and micro-robotic

injection controlled by a computer vision system uses air pressure and a glass microneedle to microinject hundreds of embryos per hour [91]. This injection is into the large yolk cell of the developing embryo and results in systemic exposure as yolk is used as the embryos' food source. Further scaling up of such micro-robotics systems is challenging, and it is difficult to integrate them with automated sorting of embryos before or after injection. This has prompted the investigation of microfluidic chips to increase micro-injection throughput. The principle of using a microfluidic device has been validated with a PDMS chip that uses electroosmotic flow to deliver pharmaceuticals by microinjection in an automated fashion, although it has not yet been scaled up to handle many embryos in parallel [92].

3.4 Advances in High-Content Imaging and Image Analysis of Zebrafish

Once embryos have been treated, the effects on embryonic development must be evaluated. For zebrafish embryos, criteria for scoring effects on the formation of various organs and tissues by morphological examination have been carefully developed, as noted earlier [3, 6, 10, 93]. The next step is to automate, quantitate, and extend these morphological examinations to gather more data from more embryos faster.

An example of basic automation and quantitation comes from a commercially available system, the IN Cell Analyzer (GE Healthcare). This device can image embryos in microtiter dishes, segment the images of embryos into the tissues that have been evaluated in manual assessments of development (eye, head, heart, muscle, etc.), and measure their area.

Advanced image capture and analysis holds the promise to provide even more detail of fine structures within organs and tissue. One such system combines high-speed optical projection tomography (which converts a series of two-dimensional images into a high-resolution three-dimensional reconstruction) with a glass capillary-based system for automated loading of embryos into the imaging chamber to image an embryo every 2.5 s [94, 95]. Subsequent processing of the images yielded data with 2.5 μm precision, allowing even subtle changes in embryonic development to be discerned. Alcian blue staining highlighted craniofacial cartilages and allowed comparisons of the effects of three different classes of teratogens. The data were accurate enough to allow construction of a dendrogram that successfully distinguished the three classes, and even different pharmaceuticals within the classes, based on the quantitative and qualitative differences on cartilage formation. Combined with techniques like automated in situ hybridization, which can label nearly any tissue by chromogenic precipitation in thousands of embryos at once [96], the optical tomography technique could be particularly helpful in evaluating the teratogenic mechanism of a pharmaceutical by comparing its effects on an organ or tissue with a database of precise 3-D information from tool pharmaceuticals with known molecular targets.

Alternatively, quantitation of changes in specific tissues can be made without having to mount or image the embryos at all. This has the advantage of eliminating the processing time, storage space, and computing power needed for images as well as the cost and technical challenges of the optical tomography system. The method, called ARQiv (automated reporter quantification in vivo) uses zebrafish expressing fluorescent proteins in a variety of tissues (Fig. 5). A fluorescent plate reader is used to rapidly capture the

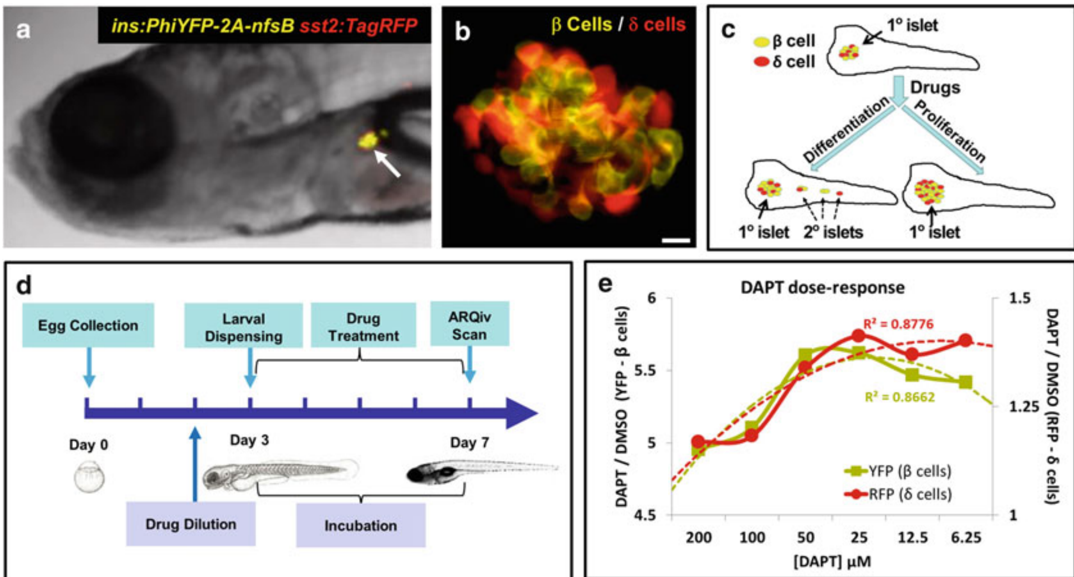

Fig. 5 Screening resources, design, and controls. (**a**) Transgenic line used for the primary screen, *Tg(ins:PhiYFP-2a-nsfB,sst2:tagRFP)lmc01* (β/δ reporter; the *insulin* promoter drives YFP-expression in β cells (*yellow*), the *somatostatin 2* promoter drives RFP expression in neighboring δ cells (*red*). Photomicrograph of the anterior region of a 7-dpf larva shows YFP and RFP labeling of the principal islet (*arrow*). (**b**) Confocal z-projection of the principal islet in a β/δ-reporter fish (scale bar: 10 μM), YFP labeling β cells (*yellow*) and RFP labeling δ cells (*red*)—note, apparent "orange" co-labeling is an artifact of z-projection in 2D format. (**c**) Illustration of two potential mechanisms by which drug exposures could lead to increased β-cell mass: (1) enhanced endocrine differentiation, indicated by secondary (2°) islet formation (*left path*) and (2) increased β-cell proliferation, indicated by supernumerary β cell numbers in the principal islet (*right path*) in the absence of effects on endocrine differentiation—that is, no effect on 2° islet formation. (**d**) Schematic of the ARQiv-HTS screening process: day 0, mass breeding produced 5000–10,000 eggs per day; day 2 (evening), JHDL compounds were serially diluted into drug plates; day 3, the COPAS-XL (Union Biometrica) was used to dispense individual 3 dpf larvae into single wells of drug plates, and plates were then maintained under standard conditions for 4 days; day 7, larvae were anesthetized and reporters quantified by automated reporter quantification in vivo (ARQiv). (E) β/δ-reporter larvae were exposed to 0.1% DMSO (negative control) or the γ-secretase/Notch inhibitor DAPT (positive control) at six different concentrations from 3 to 7 dpf. ARQiv was then used to measure fluorescent signals from β cells (*yellow line*, left *y*-axis) and δ cells (*red line*, right *y*-axis). The DAPT to DMSO ratio (DAPT/DMSO) was used to indicate signal strength for each fluorophore independently, as per the primary screen. The β-cell data show a non-monotonic dose response (*yellow dashed line*, polynomial curve fit), with maximal signal observed at 25–50 μM DAPT. The δ-cell data show a similar trend (*red dashed line*, polynomial curve fit), but with approximately fourfold lower signal strength due to higher autofluorescent background in the RFP emission range (reprinted from Wang et al. [98])

total fluorescence from the labeled tissues, revealing changes in cell number which can then be followed up with confocal imaging [97, 98]. Intriguingly, zebrafish transgenic lines that report the response of tissues to particular classes of bioactive pharmaceuticals have been developed, for instance, the estrogen response element line that expresses increasing amounts of fluorescent protein when the embryo is exposed to estrogenic pharmaceuticals [99]. By combining such transgenic lines with microtiter plate screening, the in vivo estrogenic liability of pharmaceuticals can be rapidly screened [100].

3.5 Genetics and Genomics to Elucidate Mechanism of Action Using Zebrafish

When a teratogenic liability is found for a pharmaceutical, a major challenge is to understand the mechanism of action (MoA). Is the effect on embryonic development a consequence of primary pharmacology affecting the intended target, or secondary pharmacology acting on some other molecular pathway? The zebrafish embryo provides an opportunity to unravel such questions using the many genetic and genomic tools that are available.

Two comprehensive investigations of the MoA of bioactive pharmaceuticals highlight the constellation of techniques that can be used with zebrafish, including targeted gene inactivation, genetic interaction, and RNA profiling. The first is a study of kalihinol F, a marine-sponge-derived antibiotic that causes a spectrum of defects in embryonic development, including loss of pigmentation, defective hematopoiesis and neurogenesis, and undulation of the notochord [101]. An important clue came from comparing these phenotypes to the thousands of described embryonic lethal mutations documented in zebrafish [102]. A close match was found to the *calamity* mutation, which inactivates the *atp7a* copper transporter. Indeed, upon examination of differently expressed genes by RNA sequencing of kalihinol F-treated embryos, downregulation of hemoglobin complex genes was observed, consistent with a response to decreased copper availability (copper is required to oxidize iron in hemoglobin production). Addition of extra copper directly to embryos could suppress the phenotype caused by kalinihol F treatment; conversely, addition of kalinihol F to embryos could prevent toxicity caused by an excess of copper. Thus, copper sequestration was identified as the MoA for kalinihol teratogenicity.

The second investigation uses two antiangiogenic pharmaceuticals, the natural product fumagillin and the pharmaceutical candidate analog TNP-470 [103]. Both were known to block methionine aminopeptidase 2 (MetAP-2). However, there were phenotypes in embryos treated with TNP-470 that could not be attributed to blocking angiogenesis; in particular, disruption of the convergence and extension cell movements during gastrulation.

This gastrulation phenotype was reminiscent of disrupting the noncanonical Wnt signaling pathway, but MetAP-2 had never been connected with this pathway. Using targeted knockdown of MetAP-2 in zebrafish embryos, the authors demonstrated synergy with knockdown of the noncanonical Wnt ligand, Wnt5. Moreover, they found that MetAP-2 might act upstream of calmodulin kinase II (CamKII) by rescuing MetAP-2-knockdown embryos with injection of RNA for activated CamKII, an essential clue that they exploited to dissect MetAP-2's role using cultured cells. They concluded that TNP-470 disrupted development via secondary pharmacology at a target in the noncanonical Wnt signaling pathway.

These two examples highlight some of the genomic and genetic power of the zebrafish model: a large panel of existing mutants with well-described embryonic phenotypes, a panel of antisense reagents that can be directed against any gene in the genome, and a fully curated reference genome for profiling of RNA expression [104].

A recently developed technology, Cas9/CRISPR, promises to open further vistas for zebrafish embryos in developmental toxicity [105, 106]. This technique, which allows the targeted inactivation of any gene by expression of a guide RNA and the Cas9 nuclease, avoids many of the problems with antisense reagents (including incomplete inactivation and the potential for widespread off-target effects) while retaining their speed from target identification to inactivation of a gene [107]. Moreover, inactivation of multiple genes in a single generation is possible without multi-generation breeding schemes, allowing researchers to test for effects of a pharmaceutical on more than one gene at a time [108].

4 Looking to the Horizon

The future looks exciting for alternative testing. Alternative assays can be used alone or in combination as a screening tool to predict potential for in vivo developmental toxicity, and this should be interpreted in the context of exposure. One can imagine using micro-CT to image the morphology of cultured embryos, or robotics systems for treating thousands of zebrafish embryos in microtiter plates with thousands of lead pharmaceuticals being evaluated for their teratogenic liabilities. Rapid collection and analysis of many dimensions of data on organ and tissue development from each embryo using automated image capture and analysis would be followed by comparing the phenotypes of affected embryos to databases of mutant lines (generated by CRISPR/Cas9) to evaluate potential mechanisms of action. Such information could inform the production of new lead pharmaceuticals and enable well-informed decisions on hazard assessment for other preclinical trials.

References

1. Chapin R, Augustine-Rauch K, Beyer B, Daston G, Finnell R, Flynn T, Hunter S, Mirkes P, O'Shea KS, Piersma A, Sandler D, Vanparys P, Van Maele-Fabry G (2008) State of the art in developmental toxicity screening methods and a way forward: a meeting report addressing embryonic stem cells, whole embryo culture, and zebrafish. Birth Defects Res B Dev Reprod Toxicol 83(4):446–456
2. Paquette JA, Kumpf SW, Streck RD, Thomson JJ, Chapin RE, Stedman DB (2008) Assessment of the embryonic stem cell test and application and use in the pharmaceutical industry. Birth Defects Res B Dev Reprod Toxicol 83(2):104–111
3. Brannen KC, Panzica-Kelly JM, Danberry TL, Augustine-Rauch KA (2010) Development of a zebrafish embryo teratogenicity assay and quantitative prediction model. Birth Defects Res B Dev Reprod Toxicol 89 (1):66–77
4. Barrier M, Jeffay S, Nichols HP, Chandler KJ, Hoopes MR, Slentz-Kesler K, Hunter ES 3rd (2011) Mouse embryonic stem cell adherent cell differentiation and cytotoxicity (ACDC) assay. Reprod Toxicol 31(4):383–391
5. Thomson J, Johnson K, Chapin R, Stedman D, Kumpf S, Ozolins TR (2011) Not a walk in the park: the ECVAM whole embryo culture model challenged with pharmaceuticals and attempted improvements with random forest design. Birth Defects Res B Dev Reprod Toxicol 92(2):111–121
6. Gustafson AL, Stedman DB, Ball J, Hillegass JM, Flood A, Zhang CX, Panzica-Kelly J, Cao J, Coburn A, Enright BP, Tornesi MB, Hetheridge M, Augustine-Rauch KA (2012) Inter-laboratory assessment of a harmonized zebrafish developmental toxicology assay—progress report on phase I. Reprod Toxicol 33(2):155–164
7. Selderslaghs IW, Blust R, Witters HE (2012) Feasibility study of the zebrafish assay as an alternative method to screen for developmental toxicity and embryotoxicity using a training set of 27 compounds. Reprod Toxicol 33 (2):142–154
8. Zhang C, Cao J, Kenyon JR, Panzica-Kelly JM, Gong L, Augustine-Rauch K (2012) Development of a streamlined rat whole embryo culture assay for classifying teratogenic potential of pharmaceutical compounds. Toxicol Sci 127(2):535–546
9. Panzica-Kelly JM, Brannen KC, Ma Y, Zhang CX, Flint OP, Lehman-McKeeman LD, Augustine-Rauch KA (2013) Establishment of a molecular embryonic stem cell developmental toxicity assay. Toxicol Sci 131 (2):447–457
10. Ball JS, Stedman DB, Hillegass JM, Zhang CX, Panzica-Kelly J, Coburn A, Enright BP, Tornesi B, Amouzadeh HR, Hetheridge M, Gustafson AL, Augustine-Rauch KA (2014) Fishing for teratogens: a consortium effort for a harmonized zebrafish developmental toxicology assay. Toxicol Sci 139(1):210–219
11. Panzica-Kelly JM, Zhang CX, Augustine-Rauch KA (2015) Optimization and performance assessment of the chorion-off [dechorinated] zebrafish developmental toxicity assay. Toxicol Sci 146(1):127–134
12. ICH S5(R3) Concept Paper (2015) Detection of toxicity to reproduction for medicinal products & toxicity to male fertility. www.ich.org
13. Daston GP, Beyer BK, Carney EW, Chapin RE, Friedman JM, Piersma AH, Rogers JM, Scialli AR (2014) Exposure-based validation list for developmental toxicity screening assays. Birth Defects Res B Dev Reprod Toxicol 101(6):423–428. doi:10.1002/bdrb.21132
14. New DA (1978) Whole-embryo culture and the study of mammalian embryos during organogenesis. Biol Rev Camb Philos Soc 53 (1):81–122
15. Webster WS, Brown-Woodman PD, Ritchie HE (1997) A review of the contribution of whole embryo culture to the determination of hazard and risk in teratogenicity testing. Int J Dev Biol 41(2):329–335
16. Augustine-Rauch K, Zhang CX, Panzica-Kelly JM (2010) In vitro developmental toxicology assays: a review of the state of the science of rodent and zebrafish whole embryo culture and embryonic stem cell assays. Birth Defects Res C Embryo Today 90(2):87–98
17. Lee HY, Inselman AL, Kanungo J, Hansen DK (2012) Alternative models in developmental toxicology. Syst Biol Reprod Med 58 (1):10–22
18. Kochhar DM (1980) In vitro testing of teratogenic agents using mammalian embryos. Teratog Carcinog Mutagen 1(1):63–74
19. Sadler TW, Horton WE, Warner CW (1982) Whole embryo culture: a screening technique for teratogens? Teratog Carcinog Mutagen 2 (3–4):243–253
20. Green M, Lebron J, Tanis K, Redfern B, Zhu L, Yu Y, Wang E, Kaczor A, Wysoczanski E, Chen F, Raymond C, Mattson B, Sistare F,

DeGeorge J (2016) Use of alternative developmental toxicity assays to assess teratogenicity potential of drugs and reduce animal usage. Reprod Toxicol (submitted)

21. Brown NA, Fabro S (1981) Quantitation of rat embryonic development in vitro: a morphological scoring system. Teratology 24 (1):65–78
22. Klug S, Lewandowski C, Neubert D (1985) Modification and standardization of the culture of early postimplantation embryos for toxicological studies. Arch Toxicol 58 (2):84–88
23. Van Maele-Fabry G, Delhaise F, Picard JJ (1990) Morphogenesis and quantification of the development of post-implantation mouse embryos. Toxicol In Vitro 4(2):149–156
24. Zhang CX, Danberry T, Jacobs MA, Augustine-Rauch K (2010) A dysmorphology score system for assessing embryo abnormalities in rat whole embryo culture. Birth Defects Res B Dev Reprod Toxicol 89 (6):485–492
25. Zhang C, Panzica-Kelly J, Augustine-Rauch K (2013) The rat whole embryo culture assay using the Dysmorphology Score system. Methods Mol Biol 947:423–450
26. Tonk EC, Robinson JF, Verhoef A, Theunissen PT, Pennings JL, Piersma AH (2013) Valproic acid-induced gene expression responses in rat whole embryo culture and comparison across in vitro developmental and non-developmental models. Reprod Toxicol 41:57–66
27. Naya M, Kito Y, Eto K, Deguchi T (1991) Development of rabbit whole embryo culture during organogenesis. Congenit Anom 31 (3):153–156
28. Marshall VA, Carney EW (2012) Rabbit whole embryo culture. Methods Mol Biol 889:239–252
29. Hansen JM, Carney EW, Harris C (1999) Differential alteration by thalidomide of the glutathione content of rat vs. rabbit conceptuses in vitro. Reprod Toxicol 13(6):547–554
30. Johnson JT, Hansen MS, Wu I, Healy LJ, Johnson CR, Jones GM, Capecchi MR, Keller C (2006) Virtual histology of transgenic mouse embryos for high-throughput phenotyping. PLoS Genet 2(4):e61
31. Gignac PM, Kley NJ (2014) Iodine-enhanced micro-CT imaging: methodological refinements for the study of the soft-tissue anatomy of post-embryonic vertebrates. J Exp Zool B Mol Dev Evol 322(3):166–176
32. Metscher BD (2009) MicroCT for comparative morphology: simple staining methods allow high-contrast 3D imaging of diverse non-mineralized animal tissues. BMC Physiol 9:11
33. Tomer R, Ye L, Hsueh B, Deisseroth K (2014) Advanced CLARITY for rapid and high-resolution imaging of intact tissues. Nat Protoc 9(7):1682–1697
34. Wise LD, Winkelmann CT, Dogdas B, Bagchi A (2013) Micro-computed tomography imaging and analysis in developmental biology and toxicology. Birth Defects Res C Embryo Today 99(2):71–82
35. Henning AL, Jiang MX, Yalcin HC, Butcher JT (2011) Quantitative three-dimensional imaging of live avian embryonic morphogenesis via micro-computed tomography. Dev Dyn 240(8):1949–1957
36. Gregg CL, Recknagel AK, Butcher JT (2015) Micro/nano-computed tomography technology for quantitative dynamic, multi-scale imaging of morphogenesis. Methods Mol Biol 1189:47–61
37. Chung K, Wallace J, Kim SY, Kalyanasundaram S, Andalman AS, Davidson TJ, Mirzabekov JJ, Zalocusky KA, Mattis J, Denisin AK, Pak S, Bernstein H, Ramakrishnan C, Grosenick L, Gradinaru V, Deisseroth K (2013) Structural and molecular interrogation of intact biological systems. Nature 497 (7449):332–337
38. Vickerton P, Jarvis J, Jeffery N (2013) Concentration-dependent specimen shrinkage in iodine-enhanced microCT. J Anat 223 (2):185–193
39. Wong MD, Spring S, Henkelman RM (2013) Structural stabilization of tissue for embryo phenotyping using micro-CT with iodine staining. PLoS One 8(12):e84321
40. Suter-Dick L, Alves PM, Blaauboer BJ, Bremm KD, Brito C, Coecke S, Flick B, Fowler P, Hescheler J, Ingelman-Sundberg M, Jennings P, Kelm JM, Manou I, Mistry P, Moretto A, Roth A, Stedman D, van de Water B, Beilmann M (2015) Stem cell-derived systems in toxicology assessment. Stem Cells Dev 24(11):1284–1296
41. Genschow E, Spielmann H, Scholz G, Pohl I, Seiler A, Clemann N, Bremer S, Becker K (2004) Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. Altern Lab Anim 32(3):209–244
42. Marx-Stoelting P, Adriaens E, Ahr HJ, Bremer S, Garthoff B, Gelbke HP, Piersma A, Pellizzer C, Reuter U, Rogiers V, Schenk B, Schwengberg S, Seiler A, Spielmann H, Steemans M, Stedman DB, Vanparys P, Vericat JA,

Verwei M, van der Water F, Weimer M, Schwarz M (2009) A review of the implementation of the embryonic stem cell test (EST). The report and recommendations of an ECVAM/ReProTect Workshop. Altern Lab Anim 37(3):313–328

43. Buesen R, Visan A, Genschow E, Slawik B, Spielmann H, Seiler A (2004) Trends in improving the embryonic stem cell test (EST): an overview. ALTEX 21(1):15–22
44. Höpfl G, Gassmann M, Desbaillets I (2004) Germ cell protocols. In: Differentiating embryonic stem cells into embryoid bodies, Methods in molecular biology, vol 254. Humana Press. doi:10.1385/1-59259-741-6:079
45. Chapin RE, Stedman DB (2009) Endless possibilities: stem cells and the vision for toxicology testing in the 21st century. Toxicol Sci 112(1):17–22
46. Puscheck EE, Awonuga AO, Yang Y, Jiang Z, Rappolee DA (2015) Molecular biology of the stress response in the early embryo and its stem cells. Adv Exp Med Biol 843:77–128
47. Warkus EL, Yuen AA, Lau CG, Marikawa Y (2016) Use of in vitro morphogenesis of mouse embryoid bodies to assess developmental toxicity of therapeutic drugs contraindicated in pregnancy. Toxicol Sci 149 (1):15–30
48. Marikawa Y, Tamashiro DA, Fujita TC, Alarcon VB (2009) Aggregated P19 mouse embryonal carcinoma cells as a simple in vitro model to study the molecular regulations of mesoderm formation and axial elongation morphogenesis. Genesis 47 (2):93–106
49. Lau CG, Marikawa Y (2014) Morphology-based mammalian stem cell tests reveal potential developmental toxicity of donepezil. Mol Reprod Dev 81(11):994–1008
50. Li AS, Marikawa Y (2015) An in vitro gastrulation model recapitulates the morphogenetic impact of pharmacological inhibitors of developmental signaling pathways. Mol Reprod Dev 82(12):1015–1036
51. Boxman J, Sagy N, Achanta S, Vadigepalli R, Nachman I (2016) Integrated live imaging and molecular profiling of embryoid bodies reveals a synchronized progression of early differentiation. Sci Rep 6:31623
52. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM (1998) Embryonic stem cell lines derived from human blastocysts. Science 282 (5391):1145–1147
53. Cezar GG, Quam JA, Smith AM, Rosa GJ, Piekarczyk MS, Brown JF, Gage FH, Muotri AR (2007) Identification of small molecules from human embryonic stem cells using metabolomics. Stem Cells Dev 16 (6):869–882
54. West PR, Weir AM, Smith AM, Donley EL, Cezar GG (2010) Predicting human developmental toxicity of pharmaceuticals using human embryonic stem cells and metabolomics. Toxicol Appl Pharmacol 247(1):18–27
55. Kleinstreuer NC, Smith AM, West PR, Conard KR, Fontaine BR, Weir-Hauptman AM, Palmer JA, Knudsen TB, Dix DJ, Donley EL, Cezar GG (2011) Identifying developmental toxicity pathways for a subset of ToxCast chemicals using human embryonic stem cells and metabolomics. Toxicol Appl Pharmacol 257 (1):111–121
56. Palmer JA, Smith AM, Egnash LA, Conard KR, West PR, Burrier RE, Donley EL, Kirchner FR (2013) Establishment and assessment of a new human embryonic stem cell-based biomarker assay for developmental toxicity screening. Birth Defects Res B Dev Reprod Toxicol 98(4):343–363
57. Cezar GG (2007) Can human embryonic stem cells contribute to the discovery of safer and more effective drugs? Curr Opin Chem Biol 11(4):405–409
58. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF (1995) Stages of embryonic development of the zebrafish. Dev Dyn 203(3):253–310
59. Dahm R (2002) Atlas of embryonic stages of development in the zebrafish. In: Nusslein-Volhard C, Dahm R (eds) Zebrafish: a practical approach. Oxford University Press, Oxford
60. Brannen KC, Panzica-Kelly JM, Charlap JH, Augustine-Rauch KA (2009) Zebrafish: a nonmammalian model of developmental toxicology. In: Hansen DK, Abbott BD (eds) Developmental toxicology. Target organ toxicology series, 3rd edn. Informa Healthcare USA, New York, NY, pp 215–241
61. Brannen KC, Chapin RE, Jacobs AC, Green ML (2016) Alternative models of developmental and reproductive toxicity in pharmaceutical risk assessment and the three Rs. ILAR J (in press)
62. Brand M, Granato M, Nusslein-Volhard C (2002) Keeping and raising zebrafish. In: Nusslein-Volhard C, Dahm R (eds) Zebrafish: a practical approach. Oxford University Press, Oxford

63. Brannen KC, Charlap JH, Lewis EM (2013) Zebrafish teratogenicity testing. In: Barrow PC (ed) Teratogenicity testing: methods and protocols, Methods in molecular biology, vol 947. Humana Press, New York, NY, pp 383–401
64. Truong L, Bugel SM, Chlebowski A, Usenko CY, Simonich MT, Simonich SL, Tanguay RL (2016) Optimizing multi-dimensional high throughput screening using zebrafish. Reprod Toxicol 65:139–147. doi:10.1016/j.reprotox.2016.05.015
65. Garcia GR, Noyes PD, Tanguay RL (2016) Advancements in zebrafish applications for 21st century toxicology. Pharmacol Ther 161:11–21
66. Hermsen SA, van den Brandhof EJ, van der Ven LT, Piersma AH (2011) Relative embryotoxicity of two classes of chemicals in a modified zebrafish embryotoxicity test and comparison with their in vivo potencies. Toxicol In Vitro 25(3):745–753
67. Van den Bulck K, Hill A, Mesens N, Diekman H, De Schaepdrijver L, Lammens L (2011) Zebrafish developmental toxicity assay: a fishy solution to reproductive toxicity screening, or just a red herring? Reprod Toxicol 32 (2):213–219
68. Padilla S, Corum D, Padnos B, Hunter DL, Beam A, Houck KA, Sipes N, Kleinstreuer N, Knudsen T, Dix DJ, Reif DM (2012) Zebrafish developmental screening of the ToxCast Phase I chemical library. Reprod Toxicol 33 (2):174–187
69. Yamashita A, Inada H, Chihara K, Yamada T, Deguchi J, Funabashi H (2014) Improvement of the evaluation method for teratogenicity using zebrafish embryos. J Toxicol Sci 39(3):453–464
70. Beekhuijzen M, de Koning C, Flores-Guillen ME, de Vries-Buitenweg S, Tobor-Kaplon M, van de Waart B, Emmen H (2015) From cutting edge to guideline: a first step in harmonization of the zebrafish embryotoxicity test (ZET) by describing the most optimal test conditions and morphology scoring system. Reprod Toxicol 56:64–76
71. Haldi MH, Harden M, D'Amico L, DeLise A, Seng WL (2012) Developmental toxicity assessment in zebrafish. In: McGrath P (ed) Zebrafish: methods for assessing drug safety and toxicity. Wiley, Hoboken, NJ, pp 15–25
72. Panzica-Kelly JM, Zhang CX, Augustine-Rauch K (2012) Zebrafish embryo developmental toxicology assay. Methods Mol Biol 889:25–50. doi:10.1007/978-1-61779-867-2_4
73. Henn K, Braunbeck T (2011) Dechorionation as a tool to improve the fish embryo toxicity test (FET) with the zebrafish (Danio rerio). Comp Biochem Physiol C Toxicol Pharmacol 153(1):91–98
74. Mandrell D, Truong L, Jephson C, Sarker MR, Moore A, Lang C, Simonich MT, Tanguay RL (2012) Automated zebrafish chorion removal and single embryo placement: optimizing throughput of zebrafish developmental toxicity screens. J Lab Autom 17(1):66–74
75. Brox S, Ritter AP, Kuster E, Reemtsma T (2014) Influence of the perivitelline space on the quantification of internal concentrations of chemicals in eggs of zebrafish embryos (Danio rerio). Aquat Toxicol 157:134–140
76. Gurrola-Gal MC, Penney LC, MacDougall J, Clipston AS, Brannen KC (2014) The effects of the zebrafish chorion on the response to exogenous chemicals. Birth Defects Res A Clin Mol Teratol 100(5):404
77. Nagel R (2002) DarT: the embryo test with the zebrafish Danio rerio—a general model in ecotoxicology and toxicology. ALTEX 19 (Suppl 1):38–48
78. Sipes NS, Padilla S, Knudsen TB (2011) Zebrafish: as an integrative model for twenty-first century toxicity testing. Birth Defects Res C Embryo Today 93(3):256–267
79. Wager TT, Kormos BL, Brady JT, Will Y, Aleo MD, Stedman DB, Kuhn M, Chandrasekaran RY (2013) Improving the odds of success in drug discovery: choosing the best compounds for in vivo toxicology studies. J Med Chem 56 (23):9771–9779. doi:10.1021/jm401485p
80. McGrath P, Li CQ (2008) Zebrafish: a predictive model for assessing drug-induced toxicity. Drug Discov Today 13(9–10):394–401
81. Hill AJ, Jones M, Dodd A, Diekmann H (2011) A review of developmental toxicity screening using zebrafish larvae. Int J Toxicol 30(1):105
82. Tseng HP, Hseu TH, Buhler DR, Wang WD, Hu CH (2005) Constitutive and xenobiotics-induced expression of a novel CYP3A gene from zebrafish larva. Toxicol Appl Pharmacol 205(3):247–258
83. Goldstone JV, McArthur AG, Kubota A, Zanette J, Parente T, Jonsson ME, Nelson DR, Stegeman JJ (2010) Identification and developmental expression of the full complement of cytochrome P450 genes in zebrafish. BMC Genomics 11:643
84. Jones HS, Panter GH, Hutchinson TH, Chipman JK (2010) Oxidative and conjugative xenobiotic metabolism in zebrafish larvae in vivo. Zebrafish 7(1):23–30

85. Otte JC, Schmidt AD, Hollert H, Braunbeck T (2010) Spatio-temporal development of CYP1 activity in early life-stages of zebrafish (Danio rerio). Aquat Toxicol 100(1):38–50
86. Weigt S, Huebler N, Braunbeck T, von Landenberg F, Broschard TH (2010) Zebrafish teratogenicity test with metabolic activation (mDarT): effects of phase I activation of acetaminophen on zebrafish Danio rerio embryos. Toxicology 275(1–3):36–49
87. Busquet F, Nagel R, von Landenberg F, Mueller SO, Huebler N, Broschard TH (2008) Development of a new screening assay to identify proteratogenic substances using zebrafish Danio rerio embryo combined with an exogenous mammalian metabolic activation system (mDarT). Toxicol Sci 104 (1):177–188
88. Kalueff A, Cachat J (2011) Zebrafish models in neurobehavioral research. In: Neuromethods, vol 52. Humana Press, New York, NY
89. Nishimura Y, Murakami S, Ashikawa Y, Sasagawa S, Umemoto N, Shimada Y, Tanaka T (2015) Zebrafish as a systems toxicology model for developmental neurotoxicity testing. Congenit Anom (Kyoto) 55(1):1–16
90. MacRae CA, Peterson RT (2015) Zebrafish as tools for drug discovery. Nat Rev Drug Discov. doi:10.1038/nrd4627
91. Wang W, Liu X, Gelinas D, Ciruna B, Sun Y (2007) A fully automated robotic system for microinjection of zebrafish embryos. PLoS One 2(9):e862
92. Noori A, Selvaganapathy PR, Wilson J (2009) Microinjection in a microfluidic format using flexible and compliant channels and electroosmotic dosage control. Lab Chip 9 (22):3202–3211
93. Augustine-Rauch K, Zhang CX, Panzica-Kelly JM (2016) A developmental toxicology assay platform for screening teratogenic liability of pharmaceutical compounds. Birth Defects Res B Dev Reprod Toxicol. doi:10.1002/bdrb.21168
94. Pardo-Martin C, Allalou A, Medina J, Eimon PM, Wahlby C, Fatih Yanik M (2013) High-throughput hyperdimensional vertebrate phenotyping. Nat Commun 4:1467
95. Pardo-Martin C, Chang TY, Koo BK, Gilleland CL, Wasserman SC, Yanik MF (2010) High-throughput in vivo vertebrate screening. Nat Methods 7(8):634–636
96. Coverdale LE, Burton LE, Martin CC (2008) High-throughput whole mount in situ hybridization of zebrafish embryos for analysis of tissue-specific gene expression changes after environmental perturbation. Methods Mol Biol 410:3–14
97. Walker SL, Ariga J, Mathias JR, Coothankandaswamy V, Xie X, Distel M, Koster RW, Parsons MJ, Bhalla KN, Saxena MT, Mumm JS (2012) Automated reporter quantification in vivo: high-throughput screening method for reporter-based assays in zebrafish. PLoS One 7(1):e29916
98. Wang G, Rajpurohit SK, Delaspre F, Walker SL, White DT, Ceasrine A, Kuruvilla R, Li RJ, Shim JS, Liu JO, Parsons MJ, Mumm JS (2015) First quantitative high-throughput screen in zebrafish identifies novel pathways for increasing pancreatic beta-cell mass. Elife 4:e08261
99. Gorelick DA, Halpern ME (2011) Visualization of estrogen receptor transcriptional activation in zebrafish. Endocrinology 152 (7):2690–2703
100. Gorelick DA, Pinto CL, Hao R, Bondesson M (2016) Use of reporter genes to analyze estrogen response: the transgenic zebrafish model. Methods Mol Biol 1366:315–325
101. Sandoval IT, Manos EJ, Van Wagoner RM, Delacruz RG, Edes K, Winge DR, Ireland CM, Jones DA (2013) Juxtaposition of chemical and mutation-induced developmental defects in zebrafish reveal a copper-chelating activity for kalihinol F. Chem Biol 20 (6):753–763
102. Ruzicka L, Bradford YM, Frazer K, Howe DG, Paddock H, Ramachandran S, Singer A, Toro S, Van Slyke CE, Eagle AE, Fashena D, Kalita P, Knight J, Mani P, Martin R, Moxon SA, Pich C, Schaper K, Shao X, Westerfield M (2015) ZFIN, the zebrafish model organism database: updates and new directions. Genesis 53(8):498–509
103. Zhang Y, Yeh JR, Mara A, Ju R, Hines JF, Cirone P, Griesbach HL, Schneider I, Slusarski DC, Holley SA, Crews CM (2006) A chemical and genetic approach to the mode of action of fumagillin. Chem Biol 13 (9):1001–1009
104. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W, Kilian B, Quintais LT, Guerra-Assuncao JA, Zhou Y, Gu Y, Yen J, Vogel JH, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C,

Oliver K, Clark R, Riddle C, Elliot D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthravadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper J, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Urun Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberlander M, Rudolph-Geiger S, Teucke M, Lanz C, Raddatz G, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Schuster SC, Carter NP, Harrow J, Ning Z, Herrero J, Searle SM, Enright A, Geisler R, Plasterk RH, Lee C, Westerfield M, de Jong PJ, Zon LI, Postlethwait JH, Nusslein-Volhard C, Hubbard TJ, Roest Crollius H, Rogers J, Stemple DL (2013) The zebrafish reference genome sequence and its relationship to the human genome. Nature 496(7446):498–503

105. Gonzales AP, Yeh JR (2014) Cas9-based genome editing in zebrafish. Methods Enzymol 546:377–413

106. Varshney GK, Pei W, LaFave MC, Idol J, Xu L, Gallardo V, Carrington B, Bishop K, Jones M, Li M, Harper U, Huang SC, Prakash A, Chen W, Sood R, Ledin J, Burgess SM (2015) High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. Genome Res 25(7):1030–1042

107. Kok FO, Shin M, Ni CW, Gupta A, Grosse AS, van Impel A, Kirchmaier BC, Peterson-Maduro J, Kourkoulis G, Male I, DeSantis DF, Sheppard-Tindell S, Ebarasi L, Betsholtz C, Schulte-Merker S, Wolfe SA, Lawson ND (2015) Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. Dev Cell 32(1):97–108

108. Jao LE, Wente SR, Chen W (2013) Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proc Natl Acad Sci U S A 110(34):13904–13909

Methods in Pharmacology and Toxicology (2017) 261–278
DOI 10.1007/7653_2014_27
© Springer Science+Business Media New York 2014
Published online: 03 December 2014

Using the Alternative Model *C. elegans* in Reproductive and Developmental Toxicology Studies

Daniel W. Ferreira, Yichang Chen, and Patrick Allard

Abstract

Reproduction is an extraordinarily complex biological process that requires the coordinated action of multiple cell types over the course of several months in rodent to many years in humans. A proper execution of the male and female reproductive programs is therefore crucial for the production of viable gametes and the propagation of species. Mounting evidence highlights the exquisite sensitivity of reproductive pathways to environmental influences. Therefore, there is a great need for comprehensive testing of environmental chemicals to examine their effects on reproduction. To this effect, alternative animal models, such as the nematode *Caenorhabditis elegans*, offer great advantages rooted in their biology which will be explored in this chapter. We will introduce the use of *C. elegans* in toxicology, its reproductive features that can be mobilized, and describe several validated assays that can greatly inform targeted mammalian studies.

Key words Alternative model, *C. elegans*, Toxicology, Germline, Reproduction, Apoptosis, Immunofluorescence

1 Introduction

There is a growing need for the development of accurate and reliable alternatives to traditional animal testing in toxicology. The reasons for this are ethically, economically, and scientifically based. In addition to efforts in reducing the numbers of animals used in toxicology assays, there is also interest in prioritizing a large number of compounds for toxicological assessment. Specifically, a great number of chemicals that are used routinely and in large quantities in industry remain insufficiently tested or wholly so. This need is illustrated by the toxicity testing in the twenty-first century initiative, an effort by a number of governmental and academic laboratories to test a large number of chemicals by a variety of methods including quantitative high-throughput screening assays [1].

In addition to environmental and public health concerns, there is significant economic pressure in the pharmaceutical industry to develop low-cost, high-throughput assays to predict toxic liabilities early on during drug development. Whole animal mammalian models are generally time-consuming, expensive, and require relatively large amounts of pharmaceutical material. These are

prohibitive to performing toxicological studies at the early stages of the drug development pipeline. Furthermore, these expensive studies do not necessarily predict the human response [2]. For all of these reasons, it is imperative to develop efficient, predictive and relevant alternative models in the field of toxicology.

In this chapter, we will discuss the use of alternative model systems, focusing on the utility of the nematode *Caenorhabditis elegans* as a model organism, to investigate the question of reproductive toxicity. We will first discuss areas of toxicology in which *C. elegans* has provided unique solutions to investigate toxicities at both high-throughput and mechanistic scales. We will then focus on reproductive toxicity and include the description and protocols of three assays routinely used in the field. Overall, *C. elegans* will be discussed in the context of a promising alternative model organism used to assess toxicity while reducing the use and costs of traditional whole animal mammalian models.

2 Use of *C. elegans* in Toxicology

C. elegans has been a powerful alternative model organism used by researchers over several decades. Worms are cheaply maintained in culture dishes and growth from an egg to a fertile adult is fast, approximately 72 h. The worm is small enough to use in 96- or 384-well plates for high-throughput assays yet, as a whole organism, it recapitulates many complex developmental, cellular, and physiological features that are difficult to capture in such a short time frame and/or in cell culture settings. Other advantages of using the worm are related to the powerful genetic tools available. The worm genome has been completely mapped [3] and is easily manipulated by mutation or RNAi treatment. Additionally, thousands of transgenic and mutant worm strains are available from the *Caenorhabditis* Genetics Center (CGC) for minimal cost.

There are also distinct advantages to using *C. elegans* as a model organism in toxicology assays. Treatment of worms with chemicals can be performed on either solid agar or liquid media. Assessment of toxicity is simplified by the transparency of the worm's cuticle, allowing for the observation of many endpoints without dissection of the worm itself. Simply put, the utilization of *C. elegans* allows for data to be collected in a whole, living organism at a similar scale and methodology as that employed frequently in cell line monocultures.

2.1 Genotoxicology

C. elegans has provided powerful tools to geneticists for many years. Their practicality (small size, inexpensive upkeep, self-fertilization, and quick generation time) combined with utility (fully mapped genome and a variety of available genetic manipulation techniques) in the laboratory makes worms particularly attractive to researchers

from both forward and reverse genetic approaches. DNA replication and repair machinery are well conserved in *C. elegans*, making it a relevant model organism in reproduction and genotoxicology.

In both *C. elegans* and mammals, mitochondria produce the majority of energy within a cell and are critical to normal cell homeostasis based on their roles in aging [4, 5], apoptosis [6, 7], and bioenergetics [8, 9]. Mitochondrial DNA is sensitive to both chemical and ultraviolet exposure, in part because nucleotide excision repair (NER) is absent in mtDNA (reviewed in [10]). Extensive work in this field has been done in *C. elegans* by the laboratory of Dr. Joel Meyer in describing mtDNA damage and repair mechanisms in response to genotoxicity [11–13]. When mtDNA damage persists, RNA transcription and mitochondrial function are compromised. These defects are enhanced when autophagy is inhibited, suggesting that mitophagy contributes to the overall health of the cell by removing dysfunctional mitochondria [14]. Ultraviolet exposure of human fibroblasts causes a similar induction in autophagy as in the worm, suggesting that this process is conserved [15]. Furthermore, mitophagy occurs in both mammals and worms as a result of toxicant exposure or the generation of reactive oxygen species in addition to mtDNA damage [16–18], validating the use of *C. elegans* as a model to study these mechanisms of mitochondrial dysfunction.

Although many developmental, reproductive, physiological, and cellular pathways are generally well conserved in *C. elegans* [19], it is important to recognize that other pathways are less so. For example, the metabolism of genotoxins (and other chemicals) may differ between worms and mammals as exemplified by the relative genotoxic resistance of *C. elegans* to Benzo[a]pyrene due to deficient bioactivation [11]. Thus, a certain rate of false negatives may be expected in *C. elegans* screening assays, not unlike cell culture experiments. However, metabolization differences can easily be circumvented by screening not only the parent compound but also its active metabolites.

2.2 Neurotoxicology

Worms are widely used in neurotoxicology studies for several reasons. Their sensitivity to metal-induced and mitochondrial toxicities as well as their limited but well-defined network of neurons has made them an attractive model to study neuronal dysfunction. In all, there are 302 neurons in the hermaphrodite worm, accounting for approximately 30 % of all adult cells. Major neurotransmitters are shared between humans and the worm including the serotonin, acetylcholine, gamma amino butyric acid (GABA), and dopamine (DA) systems. Thus, the worm provides a simple but conserved neuron network for investigating neurotoxicity. The short life span and transparency of *C. elegans* in combination with transgenic green fluorescent protein (GFP) tagging of neurons make the

worm an appealing model system for assessing neurodegeneration in vivo.

Neurodegeneration of DA-ergic neurons is well described in the worm, of which there are eight in the adult hermaphrodite. DA signaling is integral to worm locomotion and mechanosensitivity. Neurodegeneration of these nerves results in parkinsonian-like symptoms in *C. elegans* following exposure to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) [20], rotenone [21], manganese [22, 23], and other agents believed to promote Parkinson's disease in mammals. Similar to mammalian models, 6-hydroxydopamine-induced toxicity is selective for DA neurons in worms and is prevented by inhibiting the dopamine transporter [24], indicating that *C. elegans* is a valid alternative model for studying mechanisms of chemical-induced DA neurodegeneration.

2.3 High-Throughput Screening

C. elegans has been utilized as a model organism in several key areas of toxicology in part due to its value in high-throughput screens. *C. elegans* is a powerful organism to study gene regulation by using RNAi screens and comparative genomics. One particular such RNAi screen of 10,862 genes across three chromosomes of *C. elegans* revealed 32 target genes affecting the expression of a *Clec-85::gfp* fusion protein in these transgenic worms [25]. *Clec-85* is a C-type lectin regulated by several innate immune signaling pathways. To determine whether any of the 32 candidate genes affected mammalian immune responses, RAW264.7 mouse macrophages were treated with RNAi to each of the known mammalian orthologs. Out of the 20 orthologous genes examined, eight were found to be involved in interleukin-6 production. Thus, *C. elegans* is an effective organism in which to identify potential biological targets in vivo.

C. elegans can also be used in a high-throughput manner to investigate the mechanism of action of toxicants in vivo. RNAi of 599 genes in *C. elegans* treated with either inorganic or organic mercury revealed that knockdown of 18 genes in particular increased the susceptibility to mercury toxicity [26]. Of these, only two genes affected the toxicity of both mercurial forms, suggesting that different genes are responsible for protecting against inorganic and methyl forms. Further investigation of orthologous genes in three human cell lines confirmed differential responses in gene expression, a surprising finding since the mechanism of toxicity was previously thought to be the same for both mercurial forms based on similar toxic phenotypes—the production of oxidative stress and mitochondrial dysfunction.

Use of *C. elegans* transgenic strains can be another particularly useful tool used in high-throughput screening by exposing them to environmental contaminants in a variety of media. In this way, specific toxic mechanisms can be distinguished from complex mixtures of chemicals. For example, mutant strains deficient in

particular metabolism, detoxifying or stress response proteins, are selectively susceptible to toxicities affecting these perturbed pathways. This strategy has been used by Turner et al. in determining toxic liabilities in watershed samples collected from coal mines [27]. Taking advantage of the variety of mutant strains available in *C. elegans*, the authors narrowed down the mechanisms of observed toxicities in both stream and sediment samples by a simple screening approach. Selected mutant strains each possessed a single gene knockout rendering each sensitive to growth impairment by one of several toxic mechanisms such as oxidative stress, osmotic stress, or various metal exposures. The authors concluded that while the observed toxicity in the stream water was primarily due to osmotic stress, toxic effects elicited by sediment samples were primarily driven by the presence of metals or metalloids [27]. In this way, *C. elegans* can act as "biological sensors" when treated with environmental samples in a potentially contaminated and dangerous area.

3 *C. elegans* as a Reproductive Model

Reproduction is perhaps one of most intricate biological processes in organisms. Anomalies during reproductive processes have deep consequences on the quality of the gametes and genetic information being transferred to the next generation, thereby potentially affecting the offspring's development and physiological characteristics. Furthermore, the tight connection of reproductive pathways with development processes makes them inseparable in toxicology studies. Hence, to investigate the impairment of the reproductive function in males and females following exposure, toxicology studies need to consider processes not only taking place in adults but also during embryonic and fetal development. Faced with the complexity of biological events interconnected with reproductive process and the high number of chemicals to assay, *C. elegans* offers clear advantages relying on its reproductive features.

The reproductive system of *C. elegans* hermaphrodite consists of a symmetrically arranged bi-lobed gonad. Each lobe is U shaped, starting from the center of worm's body in opposite directions and both opening in a common uterus, itself connected to the outside through the vulva. Since the processes of meiotic differentiation, ovulation, and fertilization are spatially and temporally coupled in *C. elegans*, each and every stage of germline development can be observed at specific locations within the germline. Furthermore, all stages of meiosis can be identified through the canonical changes in nuclear architecture and morphology inherent to meiosis and are easily visualized by DAPI nuclear staining. As illustrated in Fig. 1, after mitotic proliferation while under the influence of the distal tip cells (DTC), germline nuclei will exit the mitotic zone to initiate

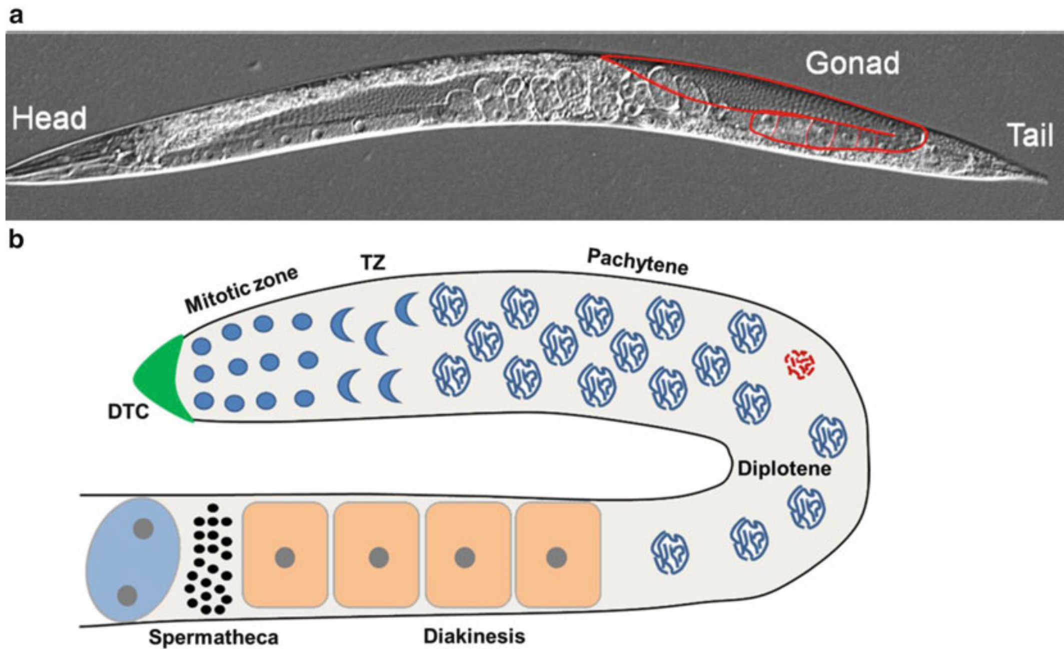

Fig. 1 Illustration of the reproductive system of *C. elegans.* (**a**) The anatomical localization of one arm of the gonad. The gonad in the tail is *circled* by the *red line*. The other gonadal arm is located on the head side of the worm and is partially hidden behind the gut. (**b**) The illustrated structure of *C. elegans*' gonad. Germline nuclei progress from the mitotic end (distal end) to the proximal end of the gonad where oocytes are fertilized. At the end of pachytene, the longest phase of meiotic Prophase I, abnormal nuclei are cleared by apoptosis (*red nucleus*). DTC = distal tip cell. TZ = transition zone (leptotene and zygotene of Prophase I)

the process of meiosis. While the nuclei display a circular morphology in mitosis, in transition zone, or leptotene and zygotene of Prophase I of meiosis, the nuclei become crescent shaped due to chromosome pairing and the initiation of synaptonemal complex formation (synapsis). At pachytene, the homologous chromosomes being intimately connected to each other, the nuclei of germ cells assume a characteristic shape where the presence of "tracks" is evident. These tracks represent the pairs of homologous chromosomes connected by a number of proteins that make up the synaptonemal complex. Several crucial meiotic events happen in this zone, namely the completion of synapsis and homologous recombination which culminates in the formation of chromosome crossovers. The proper establishment of synapsis and repair of programmed double-stranded breaks through homologous recombination are monitored in the germline and apoptotic germline nuclei are frequently observed in response to the quality control checkpoints present in late pachytene. After exiting from pachytene, germline nuclei progress through the two last remaining stages of prophase I, namely diplotene and diakinesis, where chromosomes condense further. At these stages, germline nuclei are

lined in a single file and cellularize during diakinesis to form oocytes. During this process, the synaptonemal complex disassembles concomitantly to the condensation of chromosomes such as that by the end of diakinesis, the homologous chromosomes are held together by the chiasmata, the physical product of recombination. Therefore, at diakinesis, oocytes contain six pairs of homologous chromosomes that are clearly apparent. These oocytes will move through the spermatheca, will be fertilized by sperms generated and stored earlier, and will be expelled into the worm's uterus where early embryogenesis begins. Early embryos become encapsulated by the formation of a chitinous shell and laid to the environment through the vulva as eggs.

4 Use of *C. elegans* in Reproductive Toxicology

C. elegans are particularly valuable in the laboratory as a model organism involving reproductive toxicology studies. The worm hatches from an egg, matures, and begins laying eggs of its own in approximately 72 h. This is in contrast to mice and other mammalian species in which the generation time is upwards of 10 weeks or more. During its reproductive life span of several days, a single worm will give rise to between 250 and 350 offspring through self-fertilization. Because of these high numbers, researchers commonly assess reductions in brood size as well as the growth of progeny when assessing reproductive and developmental toxicities. These endpoints are particularly attractive because they can be quantified on microtiter plates and assessed using a biosorter in medium- and high-throughput formats [28]. Thus, researchers have the opportunity to obtain data from a whole organism model in a screening fashion that is otherwise typically reserved for cell monoculture studies.

In general, the usefulness of high-throughput models is limited to their predictability in assessing toxicity in higher organisms. Boyd and colleagues assessed the accuracy of *C. elegans* in predicting mammalian toxicity following exposure to a panel of seven toxicants—three pesticides (diquat, paraquat, and parathion), two drugs (caffeine and methadone), cadmium, and the mutagen ethyl methanesulfonate (EMS). Using a biosorter, the authors established EC50 values in order to quantify the concentration of each chemical that reduced the number of offspring by 50 % over a 48 h treatment period. To assess the predictability of the assay, the worm EC50s were compared to rodent LD50 values. Parathion, cadmium, and methadone consistently ranked as the top three toxic compounds among the *C. elegans*, mouse, and rat data [28]. The most significant inconsistency was relatively minor: caffeine was the least toxic substance in the worm, which was the third and second least toxic in the mouse and rat, respectively. Overall, a strong

correlation was found between the worm and rodent studies, indicating that *C. elegans* brood size effectively predicted rodent lethality in this chemical screen. Furthermore, the data suggest that *C. elegans* represents a relevant and high-throughput alternative to the use of vertebrate animals in the assessment of chemical-induced toxicity.

As seen above, the gonad of the adult worm takes up a large part of the body mass and houses hundreds of developing germ cell nuclei. The large number of nuclei and short amount of time in which they mature illustrate the distinct time and spatial relationship of developing cells in the gonad that resembles an assembly line within a factory. One can observe nuclei in various stages of development within a single worm, an obvious advantage in assessing meiotic defects during germ cell maturation. This strategy is particularly helpful in determining mechanisms of germline toxicity that are otherwise difficult to assess in mammals due to relatively long gestational periods and low number of offspring.

One validated example of the use of *C. elegans* as an alternative model in the investigation of reproductive toxicity involves exposure to Bisphenol A (BPA). Worms exposed to BPA at doses equivalent to those used in mammalian models have increased germline nuclei apoptosis and high rates of embryonic lethality, the latter being an indication of chromosomal missegregation [29]. These events are similar to those observed in oocytes collected from BPA-treated female mice [30]. Further experiments determined that chromosome synapsis was impaired during meiosis and was associated with disruption of double-strand break repair and enhanced germline checkpoint activation in BPA-treated worms [29]. Together, the data suggest that the *C. elegans* model is a valuable alternative to traditional whole animal models in investigating germline defects and establishing mechanistic leads following chemical exposure.

5 Methods

5.1 Apoptosis Assay

Apoptosis is a genetically regulated form of cell death occurring in multicellular organisms and plays an important role in numerous biological processes including embryogenesis, aging, and the regulated process of cell division that leads to the production of gametes: meiosis. It is in these respects that it differs significantly from the cell death caused by acute cellular injury, termed necrosis.

In meiosis, cellular DNA damage and various environmental stresses can trigger genome integrity checkpoints and induce apoptosis [31], leading to germ cell loss and contributing to infertility. Hence, monitoring the number of germline nuclei undergoing apoptosis during meiosis is an important tool to investigate the

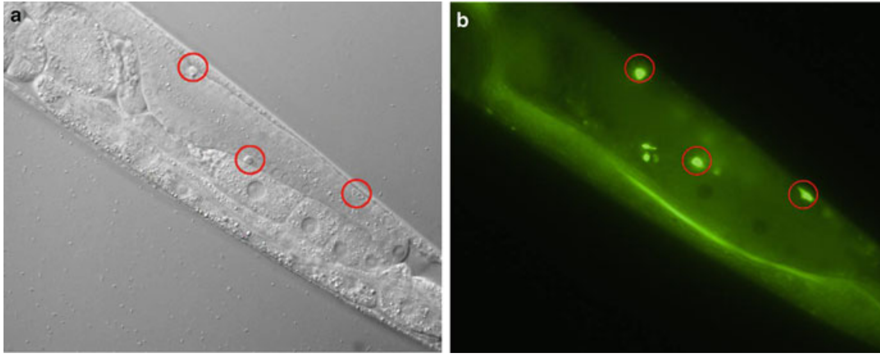

Fig. 2 Apoptotic germ cell nuclei in the pachytene zone of *C. elegans*. (**a**) The corpses of apoptotic nuclei appear in a round, bubble shape when imaged by DIC microscopy; (**b**) The apoptotic nuclei dyed with acridine orange emit an intense fluorescent signal compared to neighboring nuclei, simplifying their observation. Apoptotic nuclei are indicated by the *red circles*. The spermatheca also contains a few apoptotic cells in this example

reproductive toxicity of interested chemical or environmental agents.

Due to its transparency, *C. elegans* provides us a valuable system for observing changes in meiotic apoptosis in a living whole organism. In particular, following staining of the worm with specific dyes, the nuclei dying by apoptosis can readily be identified by fluorescence microscopy. Acridine orange (AO), a double-strand DNA dye sensitive to DNA conformation change, stains the engulfed apoptotic nucleus with a more intense fluorescence than surrounding living cells [32] (Fig. 2).

Compared with the other methods employed to monitor germ cell apoptosis in *C. elegans*, such as the identification of apoptotic corpses through DIC microscopy and use of the engulfment marker *Ced-1::gfp* strain and TUNEL assay, AO staining has several unique advantages. First, as described below, the procedure is simple and rapid. Second, AO is a cell-permeable dye and allows worms to be recovered and propagated after the assay, making it ideal for genetic screening. Finally, as there is no fixation required, the normal morphology of the cells and nuclei can be maintained.

Although AO staining provides a straightforward method to identify apoptotic nuclei in the germline, it cannot detect apoptosis in engulfment-defective strains. This caveat can be circumvented by using DIC microscopy to identify the residual corpses of apoptotic nuclei [32]. Finally, it should be noted that the total number of apoptotic nuclei can be influenced by the size of the germline. Hence, the absolute number of apoptotic germline nuclei as well as the number of apoptotic nuclei corrected by the total number of cells in pachytene should be considered.

5.1.1 Apoptosis Assay by Acridine Orange (AO) Staining

10 mg/ml AO stock solution preparation (100 ml)

1. Dissolve 1 g of AO in 100 ml of ddH_2O.
2. Vortex to dissolve and aliquot as desired.
3. Store at −20 °C and protect from light.
4. Thaw to room temperature before use and dissolve any existing precipitate by vortexing.

Staining

1. Add 399 μl M9 solution containing an ample amount of OP50 *E. coli* into a 1.5 ml Eppendorf tube.
2. Add 1 μl of AO stock solution to each tube to reach a final concentration of 25 μg/ml AO.
3. Pick 40–50 young adult worms into the tube.
4. Seal the lid with parafilm and completely cover the tube with foil to protect it from light.
5. Place the tube on a rotator for 2 h at room temperature.

Visualizing the apoptotic cells

1. Use a glass Pasteur pipette* to mix the worms to ensure that they are free of clumps of OP50 and let settle for 5 min.
2. Remove most of the supernatant from the tube and discard. Pipette remaining liquid onto the outer edges of a plate seeded with bacteria.
3. Keep the plate in the dark and let the worms recover while the plate dries (10–30 min).
4. After recovery, only pick the worms which move normally and appear healthy for use in the following steps.
5. Suspend 10–20 worms in 15 μl of M9 solution on a microscope slide and place a coverslip on it.
6. Locate the bend of the gonad in the worms with the help of the DIC filter. Switch to the Fluorescein isothiocyanate (FITC) filter to look for the bright green nuclei stained by AO.
7. In each worm, count the number of stained nuclei in both arms of the gonad.

*Note: Worms can stick to the wall of the tip if using plastic pipette tips.

5.2 Nuclear Morphology Assessment by DAPI Staining

Since the processes of meiotic differentiation, ovulation, and fertilization are spatially and temporally coupled in the *C. elegans* gonad, the dynamic chromosomal structure changes that occur during germ cell development can be observed. A single adult worm contains developing germ cells in all stages of meiosis, providing a unique system to study reproductive toxicology.

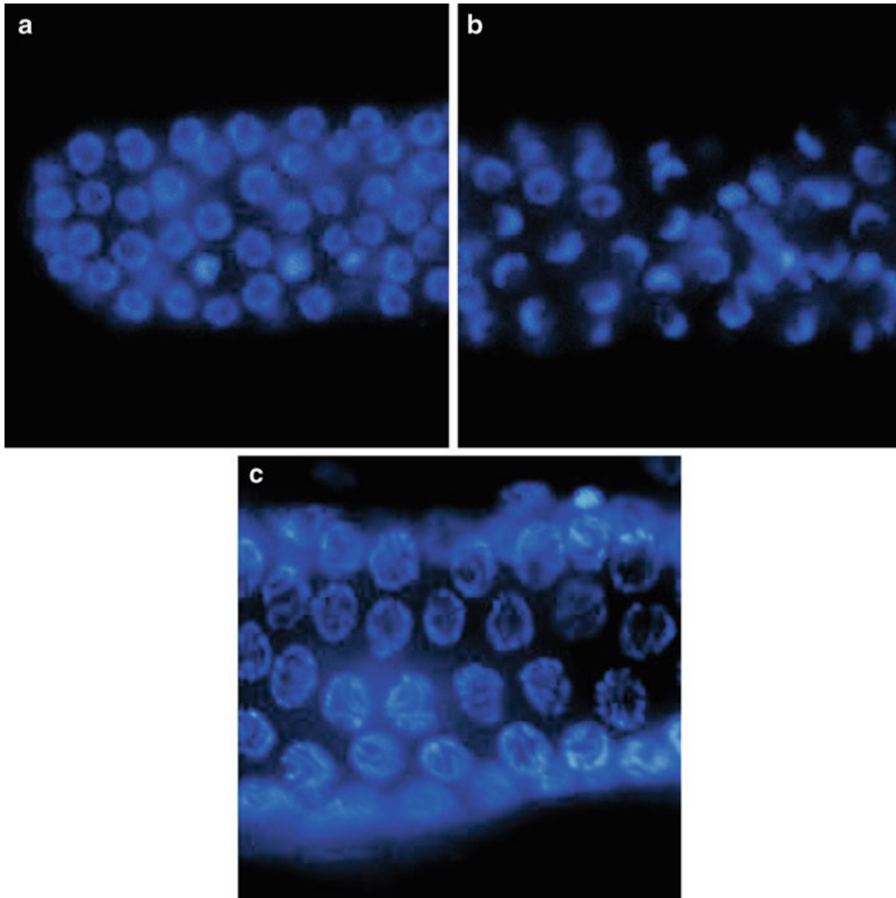

Fig. 3 Chromatin of germ cell nuclei in different meiotic stages is visualized by DAPI staining. (**a**) Mitotic zone; (**b**) Transition zone; (**c**) Pachytene zone. Notice the round shape of germline nuclei in mitotic zone; the crescent-shaped nuclei in transition zone and the presence of DNA tracks in pachytene making all these stages easily distinguishable

4′6-diamidino-2-phenylindole (DAPI), a popular nucleic acid stain, binds to dsDNA and emits a vivid blue florescence (wavelength of 461 nm) under the excitation of ultraviolet light (wavelength of 358 nm). The use of DAPI allows the visualization of DNA morphology of germline nuclei in *C. elegans* without the need for dissection, by observation through the transparent body wall (Fig. 3). The blue florescent signal is also commonly used as a counterstain during antibody staining as described in the following method.

Although the procedure of DAPI staining is quite straightforward, a fixation process is still required to allow DAPI, a cell-impermeable dye, to cross the cellular membrane and label DNA. Therefore, Carnoy's fixation, an ethanol-based fixative, is usually used prior to staining. Also of note, since the excitation wavelengths of DAPI and certain GFPs (e.g., wtGFP and GFPuv) are

very close, fluorescence overlaps can occur and the GFP signal could be quenched after an extended DAPI excitation.

5.2.1 DAPI Staining of *C. elegans*

Carnoy's fixing solution preparation (10 ml)

1. 6 ml 100 % ethanol.
2. 3 ml chloroform.
3. 1 ml acetic acid.

Mix well and store in a brown glass bottle to minimize exposure to light.

Worm fixation

1. Place a small drop (about 10 μl) of M9 solution onto a poly-lysine charged slide.
2. Pick 20–30 worms (24 h post-L4 stage) and suspend in the M9.
3. Gently absorb the excess liquid with filter paper strips. Leave a small amount of M9 to prevent drying out the worms.
4. Place two drops of Carnoy's fixative onto the worms.
5. Air-dry and properly store the slide away from light before DAPI staining.

DAPI staining

1. Add 40 μl M9 solution onto the fixed worms for rehydration.
2. Cover with an 18 × 18 mm piece of parafilm and incubate 1 h in a humidified chamber.
3. Gently remove the parafilm and absorb remaining M9 solution with filter paper strips.
4. Add 1/2 drop of DAPI solution with anti-fading agent (e.g., Fluoroshield or Vectashield) to the slide.
5. Apply a coverslip to the slide and seal with nail polish.

5.3 Examination of Protein Expression Kinetics in the Germline by Immunofluorescence

Antibody staining provides the most direct method to identify the localization and distribution of a protein of interest at both cellular and subcellular levels. Moreover, using specific antibodies allows researchers to examine posttranslational modifications of protein, such as phosphorylation or ubiquitination.

Since it has a transparent body and relatively low levels of autofluorescence (with the exception of the intestine), *C. elegans* provides the benefit of allowing the visualization of protein expression at the level of the whole organism. For example, the germline can be stained by immunofluorescence for a protein named SYP-1 which shows dynamic expression during meiosis and thereby monitors the process of synaptonemal complex formation (Fig. 4). Although using GFP or other genetically constructed strains can

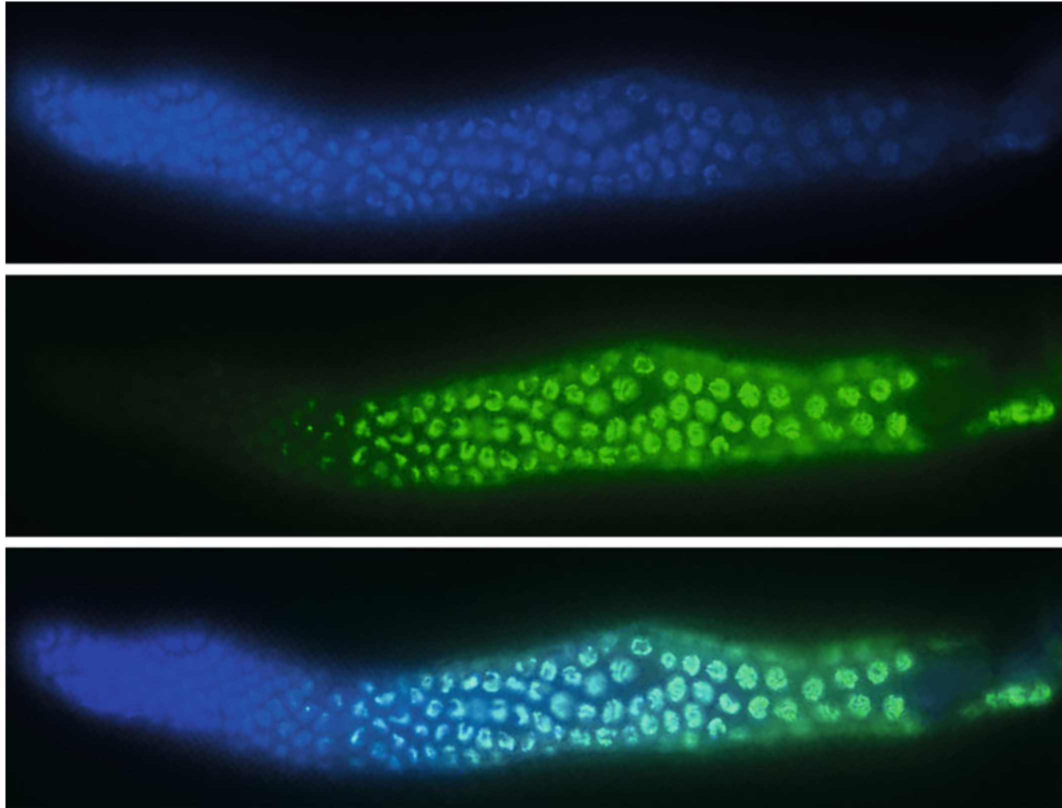

Fig. 4 Synaptonemal component protein SYP-1 expression throughout the germline of *C. elegans*. (**a**) DAPI staining of DNA; (**b**) SYP-1 staining; (**c**) Co-localization of DAPI and SYP-1. Expression of SYP-1 is evident from the beginning of meiosis (transition zone)

also provide a relatively simple way to examine protein expression, this strategy can result in abnormal regulation and expression of the expressed GFP-tagged protein [33], making antibody staining the most reliable method.

This technique comes with several challenges. Since both the shell of the embryo and the cuticle of the worm are relatively impermeable, freeze-cracking is often performed prior to staining. This allows the antibody to penetrate the target tissue more efficiently [34]. Dissecting the worms to extrude the gonad and gut can also increase antibody access to the tissue [35]. This is performed while live worms swim in a drop of M9, but can be simplified by the addition of 1 mM Levamisole to paralyze the worms. It may be helpful to carefully practice these challenging steps before performing the experiment.

5.3.1 Immunofluorescence Protocol

4 % formaldehyde fixing buffer preparation (10 ml)

1. 1 ml 10× PBS.
2. 800 μl 1 M HEPES (pH = 6.9).

3. 32 μl 0.25 M EDTA.
4. 1,080 μl 37 % liquid formaldehyde.
5. 160 μl 0.1 M $MgSO_4$.
6. 6.928 ml ddH_2O.

Prepare fresh for each experiment.

Gonad Dissection

1. Place a small drop of 1 mM Levamisole in M9 solution onto a poly-lysine charged slide (optional).
2. Pick 20–30 worms and suspend in the drop.
3. Use a surgical scalpel to cut each worm at the vulva or proximal to either the pharynx/tip of tail to extrude the gonad.
4. Fully extrude gonads by gently pipetting the solution up and down.
5. Put an 18 × 18 mm coverslip on the slide and proceed to freeze-crack.

Freeze-crack and fixation

1. At least 2 h before the experiment, prepare by placing an aluminum block and a Coplin jar filled with methanol at −80 °C.
2. After gonad dissection, place the slide on the cold aluminum block until the liquid freezes.
3. Remove the coverslip by flicking it off with a razor.
4. Immediately place the slide into cold methanol for 1 min.
5. Dry the back and sides of the slide with a tissue.
6. Quickly add 300 μl 4 % formaldehyde fixing buffer to the slide and place in a humidified chamber for 30 min.
7. Immerse the slide in PBST for 5 min and then proceed to antibody staining.

Antibody staining

1. Place slides in PBST solution with 0.5 % (w/v) BSA for at least 1 h to block nonspecific binding.
2. Add 25–50 μl of diluted primary antibody in PBST onto the slides and cover with an 18 × 18 mm piece of parafilm. Place in a humidified chamber overnight at room temperature.
3. Wash the slides three times by immersing it in PBST for 5 min.
4. Add 25–50 μl secondary antibody in PBST to the slide and cover it with parafilm. Place it in a humidified chamber for 2 h at room temperature.
5. Repeat step 3.

6. Add 1/2 drop of DAPI solution with anti-fading agent to the slide.
7. Apply a coverslip to the slide and seal.

6 Interpretation and Extrapolation of Data Generated in *C. elegans*

C. elegans diverged from the predecessors of mammalian species several hundred million years ago. Hence, as mentioned above, results obtained in the context of *C. elegans* assays should be carefully validated in other species. It is important to note that this is not specific to working with *C. elegans*, as multiple approaches and several species are now recommended to better predict human toxicity. It is also important to focus on pathways and processes that are highly evolutionarily conserved such as the ones guiding aging, neurotoxicity or, as described here, reproductive toxicity.

Ultimately, the benefits of using a tractable whole organism in the context of a high-throughput assay clearly outweigh its evolutionary distance. For example, due to a large germline and quick generation time, *C. elegans* is an efficient organism to identify toxic endpoints that typically occur at low frequency or require multigenerational experiments. One illustration of this is the detection of aneuploidies in developing oocytes. This process is simplified in *C. elegans* by observing GFP positive embryos of *Pxol-1::gfp* worms [36]. *Pxol-1* is expressed in male (XO) but not hermaphrodite (XX) worms, indicating that the presence of GFP+ embryos is the result of missegregation of the X chromosome. Importantly, this system allows for assessment of chemical-induced aneuploidy in a high-throughput fashion. Aneuploidies can also be assessed in oocytes from the ovarian follicles of mice [37]. In contrast to worms, these experiments are time-consuming and require isolation of viable oocytes from mice of a specific age, making this strategy unsuitable for a high-throughput screen.

To ensure that worm data can be properly extrapolated, it is important that *C. elegans* studies are validated by either epidemiological or mammalian experimental data. The latter approach was used to validate the *Pxol-1::gfp* worm model of chromosome missegregation. Worms were exposed to a panel of chemicals with varying mammalian reproductive toxicity and subsequently assessed for the presence of GFP + embryos. Overall, worms exposed to the selected chemicals strongly predicted mammalian reproductive toxicity following a 65 h exposure ($p = 0.008$) [36]. Thus, it may be more advantageous to utilize *C. elegans* in large-scale chemical screens and reserve the use of mice to confirm the toxicity of individual chemicals.

Data collected in *C. elegans* may also be strengthened by the inclusion of additional endpoints. The above referenced aneugenic screen was followed by an assessment of germline apoptosis following chemical treatment. As discussed earlier in this chapter, an increase in apoptosis of these nuclei is an indication of DNA checkpoint activation during meiosis [31]. Eight out of ten aneugenic compounds were successfully screened in *C. elegans* as detected by an increase of apoptotic nuclei in the gonads of treated worms [36]. By contrast, nine out of ten non-aneugenic compounds did not increase the number of apoptotic nuclei. Together, the combination of the apoptosis assay and the reporter-based high-throughput assay gives strong evidence that compounds that are aneugenic in mammals can be successfully screened by *C. elegans* in an accurate fashion that is faster than can be performed in whole animal mammalian models.

7 Concluding Remarks

Key pathways are conserved in the worm including much of the DNA replication and repair machinery, as well as mitochondrial function. It is for these reasons that *C. elegans* is particularly popular with researchers in the fields of aging, genetics, and reproduction. Toxicologists are increasingly using worms in the laboratory because of their broad utility and the relative quickness with which experiments can be performed. There are distinct advantages to using *C. elegans* in biological sciences due to the ease and inexpensive nature of worm culture, fast generation time, transparent cuticle, and the multitude of genetic tools available.

Though *C. elegans* is an appealing model for use in toxicology, the importance of validating worm data with mammalian studies must be stressed. With this in mind however, there are clear benefits that *C. elegans* affords that cannot be otherwise gained by using traditional whole animal models. One of the greatest benefits of using the worm is its value in high-throughput screens. Worms can be screened following exposure to various chemicals by endpoints such as survival, growth, and fluorescence in multiwell formats, giving researchers the opportunity to examine a whole organism by methods typically reserved for cell monoculture studies. This is extremely important when considering the number of chemicals in the environment and in industrial settings that have not undergone sufficient toxicity testing. In conclusion, the use of *C. elegans* falls in line with the mandate of the existing Tox21 initiative and its use as an alternative model has already yielded a positive impact on the field of toxicology.

References

1. National Research Council (2007) Toxicity testing in the 21st century: a vision and a strategy. The National Academies Press, Washington
2. Olson H et al (2000) Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul Toxicol Pharmacol 32(1):56–67
3. C. elegans Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans:* a platform for investigating biology. Science 282(5396):2012–2018
4. Trifunovic A et al (2004) Premature ageing in mice expressing defective mitochondrial DNA polymerase. Nature 429(6990):417–423
5. Braeckman BP et al (2002) Assaying metabolic activity in ageing *Caenorhabditis elegans.* Mech Ageing Dev 123(2–3):105–119
6. Tann AW et al (2011) Apoptosis induced by persistent single-strand breaks in mitochondrial genome: critical role of EXOG (5'-EXO/endonuclease) in their repair. J Biol Chem 286(37):31975–31983
7. Senoo-Matsuda N et al (2003) A complex II defect affects mitochondrial structure, leading to ced-3- and ced-4-dependent apoptosis and aging. J Biol Chem 278(24):22031–22036
8. Ishiguro H et al (2001) Enhancement of oxidative damage to cultured cells and *Caenorhabditis elegans* by mitochondrial electron transport inhibitors. IUBMB Life 51(4):263–268
9. Li N et al (2003) Mitochondrial complex I inhibitor rotenone induces apoptosis through enhancing mitochondrial reactive oxygen species production. J Biol Chem 278 (10):8516–8525
10. Meyer JN et al (2013) Mitochondria as a target of environmental toxicants. Toxicol Sci 134 (1):1–17
11. Leung MC et al (2010) *Caenorhabditis elegans* generates biologically relevant levels of genotoxic metabolites from aflatoxin B1 but not benzo[a]pyrene in vivo. Toxicol Sci 118 (2):444–453
12. Boyd WA et al (2010) Nucleotide excision repair genes are expressed at low levels and are not detectably inducible in *Caenorhabditis elegans* somatic tissues, but their function is required for normal adult life after UVC exposure. Mutat Res 683(1–2):57–67
13. Meyer JN et al (2007) Decline of nucleotide excision repair capacity in aging *Caenorhabditis elegans.* Genome Biol 8(5):R70
14. Bess AS et al (2012) Mitochondrial dynamics and autophagy aid in removal of persistent mitochondrial DNA damage in *Caenorhabditis elegans.* Nucleic Acids Res 40(16):7916–7931
15. Bess AS et al (2013) UVC-induced mitochondrial degradation via autophagy correlates with mtDNA damage removal in primary human fibroblasts. J Biochem Mol Toxicol 27(1):28–41
16. Gong K et al (2012) Autophagy-related gene 7 (ATG7) and reactive oxygen species/extracellular signal-regulated kinase regulate tetrandrine-induced autophagy in human hepatocellular carcinoma. J Biol Chem 287 (42):35576–35588
17. Bodhicharla R et al (2014) The tobacco-specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) induces mitochondrial and nuclear DNA damage in *Caenorhabditis elegans.* Environ Mol Mutagen 55(1):43–50
18. Tam ZY et al (2014) Effects of lithium on age-related decline in mitochondrial turnover and function in *Caenorhabditis elegans.* J Gerontol A Biol Sci Med Sci 69(7):810–820
19. O'Neil N, Rose A (2006) DNA repair. WormBook. p. 1–12
20. Pu P, Le W (2008) Dopamine neuron degeneration induced by MPP+ is independent of CED-4 pathway in *Caenorhabditis elegans.* Cell Res 18(9):978–981
21. Zhou S, Wang Z, Klaunig JE (2013) *Caenorhabditis elegans* neuron degeneration and mitochondrial suppression caused by selected environmental chemicals. Int J Biochem Mol Biol 4(4):191–200
22. Benedetto A et al (2010) Extracellular dopamine potentiates mn-induced oxidative stress, lifespan reduction, and dopaminergic neurodegeneration in a BLI-3-dependent manner in *Caenorhabditis elegans.* PLoS Genet 6(8):1–18
23. Settivari R et al (2013) The Nrf2/SKN-1-dependent glutathione S-transferase pi homologue GST-1 inhibits dopamine neuron degeneration in a *Caenorhabditis elegans* model of manganism. Neurotoxicology 38:51–60
24. Nass R et al (2002) Neurotoxin-induced degeneration of dopamine neurons in *Caenorhabditis elegans.* Proc Natl Acad Sci U S A 99(5):3264–3269
25. De Arras L et al (2013) Comparative genomics RNAi screen identifies Eftud2 as a novel regulator of innate immunity. Genetics 197 (2):485–496
26. McElwee MK et al (2013) Comparative toxicogenomic responses of mercuric and methylmercury. BMC Genomics 14:698
27. Turner EA et al (2013) Assessing different mechanisms of toxicity in mountaintop

removal/valley fill coal mining-affected watershed samples using *Caenorhabditis elegans*. PLoS One 8(9):e75329

28. Boyd WA et al (2010) A high-throughput method for assessing chemical toxicity using a *Caenorhabditis elegans* reproduction assay. Toxicol Appl Pharmacol 245(2):153–159
29. Allard P, Colaiacovo MP (2010) Bisphenol A impairs the double-strand break repair machinery in the germline and causes chromosome abnormalities. Proc Natl Acad Sci U S A 107 (47):20405–20410
30. Hunt PA et al (2003) Bisphenol a exposure causes meiotic aneuploidy in the female mouse. Curr Biol 13(7):546–553
31. Gartner A et al (2000) A conserved checkpoint pathway mediates DNA damage–induced apoptosis and cell cycle arrest in *C. elegans*. Mol Cell 5(3):435–443
32. Gumienny TL et al (1999) Genetic control of programmed cell death in the *Caenorhabditis elegans* hermaphrodite germline. Development 126(5):1011–1022
33. Mello C, Fire A (1995) DNA transformation. Methods Cell Biol 48:451–482
34. Strome S, Wood WB (1982) Immunofluorescence visualization of germ-line-specific cytoplasmic granules in embryos, larvae, and adults of *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 79(5):1558–1562
35. Crittenden SL et al (1994) GLP-1 is localized to the mitotic region of the *C. elegans* germ line. Development 120(10):2901–2911
36. Allard P et al (2013) A *C. elegans* screening platform for the rapid assessment of chemical disruption of germline function. Environ Health Perspect 121(6):717–724
37. Hodges CA et al (2002) Experimental evidence that changes in oocyte growth influence meiotic chromosome segregation. Hum Reprod 17(5):1171–1180

Methods in Pharmacology and Toxicology (2017) 279–312
DOI 10.1007/7653_2016_66
© Springer Science+Business Media New York 2016
Published online: 08 October 2016

High-Throughput Screens for Embryonic Stem Cells: Stress-Forced Potency-Stemness Loss Enables Toxicological Assays

Quanwen Li, Yu Yang, Erica Louden, Elizabeth E. Puscheck, and Daniel A. Rappolee

Abstract

Embryonic stem cells (ESCs) respond to the majority of toxicological stresses by growing more slowly, and a subpopulation of cells loses all of a subset of nuclear potency/stemness factors. Slowed growth and potency factor loss occur despite culture of ESCs under conditions that favor proliferation and potency maintenance. Stemness/potency factor loss enables the stem cells to differentiate to create essential first lineage functions. Increase in the fraction of stem cells with potency loss and differentiation increase, in response to diminished population growth, is called compensatory differentiation since increased differentiation compensates for diminished population growth.

To take advantage of compensatory differentiation where the potency marker Rex1 undergoes permanent stress-induced loss (Slater et al., Stem Cells Dev 23:3049–3064, 2014), a high throughput screen (HTS1 ESC) has been created in ESCs. Using low-stress lentivirus infection and low-stress FACS isolation, ESC lines were created that are transgenic for gene constructs that use either or both Rex1 potency factor promoter to drive red fluorescence protein (Rex1-RFP) and/or Oct4 potency factor promoter to drive green fluorescent (Oct4-GFP) (Li et al., Stem Cells Dev 25:320–328, 2016). Four independent assays showed that control hyperosmotic stress caused Rex1-RFP to decrease. These assays were immunofluorescence, microplate reader, flow cytometry, and Western blot. In addition, Hoechst was used to assay cell number at the start and end of treatment in order to quantitate growth rate. Since potency factor loss enables differentiation increase then stress-forced differentiation should compensate for diminished stem cell population size. Stress dose-dependent induction of decreased Rex1-RFP was similar to, and corroborated, by dose-dependent loss of endogenous Rex1 protein. This assay shows promise in reporting toxicological stresses.

Potency loss using Rex1-RFP has been quantitated and at a nonmorbid stress dose approaches a 20 % increase over unstressed ESCs. The quantitation of the gain of differentiation in subpopulation size is underway [Li et al., Development (manuscript in preparation)], and hypothetically is within the 20 % of cells with complete (e.g., to the level of non-transgenic parental ESCs) Rex1-RFP loss. Four days after fertilization the ESC lineage arises in the blastocyst stage embryo. At the same time the distinct placental trophoblast stem cell lineage (TSC) also arise and these cells undergo a 50 % induction of giant cell differentiation as measured by morphology, and gain of first lineage differentiated markers. For both ESC and TSC, stress-forced potency loss and differentiation increase occurs despite culture conditions that should maintain proliferation and potency. Thus these assays measure stem cells depletion due to stress forced growth and potency decrease, and differentiation increase. Production of a TSC HTS is important as this stress-forced differentiation is larger than ESC effects and is linked to events that are known to lead to miscarriage.

Keywords: Developmental and reproductive toxicology, High throughput screen (HTS), Embryonic stem cells (ESC), Placental trophoblast stem cells (TSC), Compensatory differentiation, Potency factor

1 Introduction

Within 5 days after fertilization human and rodent embryos allocate embryonic stem cells (ESCs) and placental trophoblast stem cell lineages (TSCs) [1]. At this time the embryo enters the unique phase of the mammalian life history of exponential growth when nearly all cells proliferate and the first lineage differentiates in TSCs and ESCs to produce nutrition-acquiring function [1, 2]. This period continues through embryo implantation into the uterus, persists through gastrulation and lasts for weeks in first trimester human pregnancy.

The strategy of the early embryo is to expand rapidly in size and differentiate subpopulations of cells to carry out minimal essential functions for embryo survival. This is illustrated well by the need for first placental lineage antiluteolytic hormones to increase rapidly after implantation or the embryo undergoes menstrual loss [3, 4]. Endocrine antiluteolytic hormones from the implanting embryo sustain the corpus luteum of the ovary or it degenerates. In turn the corpus luteum produces endocrine hormone progesterone that returns to the implantation site and induces the uterine glandular epithelium to continue secreting nutrients apposite the rapidly growing embryo. It is clear that the first differentiated lineage from the ESCs in the embryo also acquires nutrients [5, 6]. This is the superficial extraembryonic yolk sac endoderm (XEN) which is necessary to sustain [5] the adjacent, underlying ESC lineage. We think that for either ESCs or TSCs, stress induces runting—diminished stem cell population expansion—that is compensated by forced differentiation to create minimal essential nutrition acquisition mediated by first differentiated lineage cells.

Many toxicological and other stressors slow growth of the embryo during the period of exponential growth. Paradoxically the smaller-than-normal stem cell population further reduces stem cell number by forcing nuclear potency/stemness factor loss and induction of differentiation to first lineage in ESCs [7] and TSCs [8–11]. This forced differentiation occurs in the presence of conditions that should maintain potency, but stress blocks potency to create extraembryonic lineages that acquire nutrition for the rapidly growing embryo [12, 13]. If the stress-induced runting and forced differentiation is too great it is likely that the embryo is lost and miscarries due to insufficient first lineage function. These two stress-induced responses also occur in TSCs [9–17], but in TSCs a third phase of the response has been characterized and this is irreversible differentiation of over half of the stem cells [14]. If all three steps occur in the embryo in vivo, this would likely lead to miscarriage.

In a previous study it was shown that robust, nonmorbid levels of stress in ESCs cause loss of four nuclear potency factors, Oct4, Sox2, Nanog, and Rex1 by 4 h of stress, but that all, save Rex1, had rebounded to unstressed baseline levels by 24 h [7]. Although the ESCs are cultured using conditions that normally maintain high levels of proliferation and potency, nonmorbid stress forces decreased proliferation and potency/stemness factor loss despite these conditions. Thus for both lineages in the early embryo, stem cells are lost in two ways, by decreased growth and increased differentiation.

To create a high throughput screen (HTS), ESCs were made transgenic for Rex1-RFP, Oct4-GFP, or both [18]. It was shown by immunofluorescence, microplate reader, flow cytometry, and immunoblot that as cell growth diminishes the number of stem cells with potency loss increases in proportion to increasing stress. As in the previous study Rex1-RFP (aka ESC HTS1) reports a chronic loss (i.e., after three days of stress) of potency but Oct4-GFP reporter does not. Early tests of larger toxicant sets, such as a 17 toxicant subset of a set previously used in an ECVAM (European Committee for Validation of Alternate Methods) stem cell validation [19], show promise in ESC HTS1 (*see* Section 3.5.3). It is the ESC HTS1 using Rex1-RFP that we describe here.

It should be noted that other toxicological tests have been developed using ESCs; however, these tests remove conditions (e.g., the growth factor LIF) during cell culture that promote proliferation and stemness/potency for 1–3 weeks prior to adding toxicants. Thus, these tests predifferentiate ESCs to feature high subpopulations of cardiomyocytes (i.e., or hepatocytes or neurons) and are thus not really performing a stem cell test but a test on developmental cardiotoxicity [20–22] or for other differentiated cell types. Organ-specific developmental toxicology is an important application of ESC technology. But naming these toxicity assays ESC assays is about the same as naming in vivo D.A.R.T. (Developmental and Reproductive Toxicology) assays "lab chow" assays. Lab chow is permissive to female lab animals supporting pregnancy and toxicologists testing, as ESCs are permissive to enabling differentiation to many cell types and testing toxicity on those cells. But much current use of ESC for high throughput screens is actually a testing of differentiated parenchymal cells from the ESCs, but not the ESCs themselves.

Thus, there are several important advantages in testing ESCs maintained as ESCs. One advantage is that nonmorbid stress may leave stress memories in ESCs and these cells then differentiate into primordial germ cells and then into oocytes and sperm that carry the toxicological stress via epigenetic memories into multiple future generations. ESCs maintained as ESCs during stress testing go on to make primordial germ cells that produce oocytes and sperm as well as all the other parenchymal lineages in the adult. A second advantage is that ESCs (as well as TSCs) appear to prioritize differentiation under stress. This means that when stress forces potency

loss, resulting differentiation is to first lineage at the deficit of later lineages [12]. Although not fully proven in vitro, and awaiting in vivo tests, later lineages such as cardiomyocytes may also be decreased indirectly by directly stressing ESCs. Finally a third advantage is that ESCs may have an expected stress response; with prolonged stress irreversible differentiation occurs. Both hyperosmotic [10] and hypoxic stress [14], leads to irreversible TSC differentiation with either high stress level or long duration. If irreversibility is a property of stressed ESCs, this phenomenon might lead to miscarriage. Thus, HTSs in TSCs and ESCs may ultimately test toxicological exposures that cause miscarriage. We describe here the patented assay using Rex1-RFP (HTS1 ESC0) but not other patented assays for a first ESC lineage promoter or first lineage TSC promoter driving GFP (HTS2 for ESC and TSC, respectively).

2 Materials

2.1 Media and Solutions

2.1.1 Dulbecco's High Glucose Modified Eagles Medium (We Use HyClone®)

Glycine, 30 mg/L

L-arginine hydrochloride, 84 mg/L

L-cystine 2 HCl, 63 mg/L

L-glutamine, 584 mg/L

L-histidine hydrochloride-H2O, 42 mg/L

L-isoleucine, 105 mg/L

L-leucine, 105 mg/L

L-lysine hydrochloride, 146 mg/L

L-methionine, 66 L-phenylalanine, 30 mg/L

L-serine, 42 mg/L

L-threonine, 95 mg/L

L-tryptophan, 16 mg/L

L-tyrosine disodium salt dehydrate, 104 mg/L

L-valine, 94 mg/L

Choline chloride, 4 mg/L

D-calcium pantothenate, 4 mg/L

Folic acid, 4 mg/L

Niacinamide, 4 mg/L

Pyridoxine hydrochloride, 4 mg/L

Riboflavin, 0.4 mg/L

Thiamine hydrochloride, 4 mg/L

i-inositol, 7.2 mg/L

Calcium chloride ($CaCl_2$) (anhyd.), 200 mg/L

Ferric nitrate ($Fe(NO_3)_3 9H_2O$), 0.1 mg/L

Magnesium sulfate ($MgSO_4$) (anhyd.), 97.67 mg/L
Potassium chloride (KCl), 400 mg/L
Sodium bicarbonate ($NaHCO_3$), 3700 mg/L
Sodium chloride (NaCl), 6400 mg/L
Sodium phosphate monobasic (NaH_2PO_4-H_2O), 125 mg/L
D-glucose (Dextrose), 4500 mg/L
Phenol red, 15 mg/L

2.1.2 MEM Nonessential Amino Acid Solution 100× (We Use Sigma®)

L-alanine (free base), 890 mg/L
L-asparagine H_2O, 1500 mg/L
L-aspartic acid, 1330 mg/L
L-glutamic acid, 1470 mg/L
Glycine, 750 mg/L
L-proline, 1150 mg/L
L-serine, 1050 mg/L

2.1.3 ESC Growth Medium

Dulbecco's High Glucose Modified Eagles Medium
Supplemented with 1× MEM nonessential amino acid solution
L-glutamine, 1 mM
Sodium pyruvate, 1 mM
β-mercaptoethanol, 7 μL
Penicillin, 100 unit/mL
Streptomycin, 100 μg/mL
Millipore® ESC-qualified EmbryoMax fetal bovine serum, 150 mL/L
Millipore® ESC-qualified ESGRO® Leukemia Inhibitory Factor (LIF), 1000 unit/mL

2.1.4 Dulbecco's Phosphate Buffered Saline (We Use HyClone®)

Potassium chloride (KCl), 200 mg/L
Potassium phosphate monobasic (KH_2PO_4), 200 mg/L
Sodium chloride (NaCl), 8000 mg/L
Sodium phosphate dibasic (Na_2HPO_4-$7H_2O$), 2160 mg/L, pH 7.4.

2.1.5 Dulbecco's Phosphate Buffered Saline with Calcium and Magnesium (We Use HyClone®)

Calcium chloride ($CaCl_2$) (anhyd.), 100 mg/L
Magnesium chloride ($MgCl_2$-$6H_2O$), 100 mg/L
Potassium chloride (KCl), 200 mg/L
Potassium phosphate monobasic (KH_2PO_4), 200 mg/L
Sodium chloride (NaCl), 8000 mg/L
Sodium phosphate dibasic (Na_2HPO_4-$7H_2O$), 2160 mg/L, pH 7.4.

2.1.6 Gelatin Solution

The EmbryoMax 0.1 % gelatin solution has been optimized and validated for Stem cell culture & is available in a 500 ml format (EMD Millipore, Billerica, Massachusetts).

2.2 Cell Culture Ware

Cells are routinely cultured in 100 ml or 60 ml tissue culture treated dish (Santa Cruz Biotech, Dallas, TX). For stress treatment, cells are cultured on 6-well plates, black 96-well flat bottom plates or clear round bottom 96-well plates (Costar, Corning, NY). All cell culture plastics are precoated with 0.1 % gelatin solution for 30 min at room temperature.

2.3 Cell Lines

Germline competent mESC-D3 cells were purchased from ATCC (Manassas, VA). mESC-D3 ESCs [23].

2.4 Lentiviral Particles and Cell Infection

Rex1 promoter reporter lentiviral particles that express the red fluorescent protein mApple were from Allele Biotechnology (San Diego, CA). The Rex1-RFP reporter construct uses the mouse 0.7-kb minimal response promoter for F9 teratocarcinoma, used in human ESC to drive GFP [24]. Oct4 promoter reporter lentiviral particles that express the green fluorescent protein copGFP and TranDux™ transduction reagent were from System Biosciences (Mountain View, CA). The Oct4-GFP reporter construct replacing the mouse 2.3-kb minimal promoter for ESC and mouse blastocyst ICM was defined previously [18, 25]. Virus infection of ESCs followed the manufacturers' instructions. These may include either beta-lactamase or luciferase [26, 27], instead of GFP or RFP.

2.5 Additional Reagents

Hoechst, Invitrogen (Grand Island, NY)

BCA protein assay reagents were from Thermo Scientific (Rockford, IL).

Positive control for forced potency loss: Sorbitol, Sigma, in culture medium

2.6 ECVAM ESC Test [19, 28, 29] Embryotoxic Stressors

2.6.1 Strongly Embryotoxic Group

Hydroxyurea, Sigma, in DMSO

6-aminonicotinamide, Sigma, in DMSO

3-bromo-2-deoxyuridine, Sigma, in DMSO

Methotrexate, Sigma, in DMSO

Methylmercury chloride, Sigma, in DMSO

All trans retinoic acid, Sigma, in DMSO

2.6.2 Moderately Embryotoxic Group

Boric acid, Sigma, in DMSO

Salicylic acid sodium salt, Sigma, in water

Lithium chloride, Sigma, in water

Dimethadione, Sigma, in DMSO

Valproic acid (VPA), Sigma, in DMSO

Methoxyacetic acid, Sigma, in water

2.6.3 Nonembryotoxic Group

Acrylamide, Sigma, in water

Dimethyl phthalate, Sigma, in DMSO

D-(+)-camphor, Sigma, in DMSO

Diphenhydramine hydrochloride, Sigma, in water

Penicillin-G sodium salt, Sigma, in water

2.6.4 AMPK-Modulating Diet Supplements Group

Coenzyme Q10 (CoQ10), Sigma, in DMF

BioResponse 3,3′-diindolylmethane (BR-DIM), BioResponse Nutrients, LLC., in DMSO

Metformin, CVS/pharmacy, in water

Sodium salicylate, Sigma, in water

Resveratrol, (Resvitale), in DMSO

3 Methods

3.1 Generation of Mouse Transgenic ESCs with Single- or Double-Viable Potency Activity Reporters by Low Stress Protocols

When possible, our goal is to create reporter stem cells using protocols that maintain a low stress history. This is done by avoiding (1) electroporation or cationic lipids that kills or stresses a majority of cells and instead uses receptor-mediated lentivirus infection and (2) selection by antibiotic resistance which creates crisis events and instead we use FACS to select cells expressing potency reporters without death [18]. The starting confluence of cells is 20 % or 100,000 cells per well to allow most of the cells to proliferate after virus infection, as proliferation is required for the virus mediated transgene to integrate into the cellular genome. The cells are incubated in 37 °C for 2–3 h for attachment and readiness for virus infection. A virus infection medium is made by mixing Oct4 and/or Rex1 promoter reporter lentiviral particles into a regular growth medium supplemented with 1× TransDux. The infection medium is applied to mESCs at 400 mL per well. After infection, the cells are cultured for 3 days to allow the expression of the fluorescent reporter proteins. After expansion of the total cell number by passaging, the infected cells are subjected to FACS sorting using a BD FACS Vantage SE cell sorter (BD Biosciences, San Jose, CA) to obtain pure fluorescent cells. Heterogeneity may reestablish after culturing for too many passages. In this case, the normally small population of fluorescent dim cells should be monitored and evaluated under fluorescent microscope. The best solution of this problem is to use cells that have small passage numbers. Reporter ESCs can also be subject to re-FACS to isolate and expand to regain nearly quantitative expression of viable reporter fluorescence transgenes.

3.2 Transgenic ESC Culture and Test Drug or Toxicant Treatment

Transgenic germline-competent mESC-D3 cells are cultured in the absence of feeder cells in DMEM supplemented with 15 % mESC-screened fetal bovine serum, 2 mM L-glutamine, 1 mM sodium pyruvate, 1 mM nonessential amino acids, 0.1 mM 2-mercaptoethanol, and 1000 U/mL murine LIF on 0.1 % gelatin-coated dishes at 37 °C in humidified air with 5 % CO_2. mESCs are passaged onto 96-well plates and cultured overnight after passaging before stimulation with drugs. The standard time of drug treatment is 3 days. Osmolality of ESC media and LIF with added 200 mM sorbitol is used as the standard positive control of stress-forced potency loss. ESCs cultured with the normal growth medium with LIF but without drug are used as the standard negative control for normal potency and proliferation maintenance.

In addition to the above 3-day controls, a time zero control is also included by setting aside a plate with the identical amount of cells as in the 3-day stress plates. These cells on the time zero plate are assayed and terminated from culture at the start of toxicant exposure or control stresses. Then the time zero cells are determined for unstressed potency reporter and for Hoechst-stained nuclei. The time zero reporter signal is used to evaluate the starting potency level. The time zero Hoechst-stained nuclei signal is used to evaluate the changes of cell growth rate under different drugs as after fluorescence reading at time final. The doubling rate from time final—time zero is then determined for each dose.

3.3 Determination of Potency Reporter by Fluorescent Plate Reading

After toxicant treatment, the cells are briefly rinsed with DPBS with calcium and magnesium, 100 μL per well, aspirated and another 100 μL of DPBS with calcium and magnesium is added per well and the plate is ready for plate reading. The excitation and detection peak wavelengths for Rex1-mApple are 568 nm and 611 nm, respectively. The optimal day for measurement of Rex1-RFP is identified as 3 days by a time course experiment in which 3 day time point showed the substantial reduction of fluorescence without significant change of the cellular phenotype or increased morbidity (for details, refer to "Notes" (Section 4) below, also refer to [18]). The excitation and detection of peak wavelengths for Oct4-copGFP are 485 nm and 528 nm, respectively. After live-cell reading, the cells are stained with 5 μM of Hoechst 33342 in DPBS with calcium and magnesium, 50 μL per well, for 30 min at 37 °C incubator. The cells are briefly rinsed with DPBS with calcium and magnesium, 100 μL per well, aspirated and another 100 μL of DPBS with calcium and magnesium is added per well and the plate is ready for plate reading of Hoechst stained nuclei. The excitation and detection peak wavelengths for Hoechst 33342 stained nuclei are 392 nm and 440 nm, respectively. The time zero plate of cells are determined for potency reporter and for Hoechst stained nuclei.

3.4 Fluorescence Data Analysis

The raw reading of the red (Rex1-RFP) or green (Pdgfra-GFP) reporter fluorescence is normalized to the blue fluorescence of the Hoechst-stained nuclei from the cells. The Pdgfra-GFP ESC line was a kind gift from Dr Kat Hadjantonakis and Jerome Artus [30–32] who previously characterized this GFP knocked into the ESC first lineage marker Pdgfra. We have adapted this to complement assays for stress-forced Rex1-RFP or Rex1 endogenous protein loss [7, 18], and FGF5 decrease [7], which together indicate that Pdgfra/first lineage should occur.

This red–blue ratio represents the average potency level per cell or green/blue value represents the average first lineage differentiation level per cell. Any decrease of the red/blue or increase of the green/blue value reflects the potency loss or differentiation gain under cellular stress treatment. To compare the results from different independent biological experiments, the red–blue ratio or green/blue values from drug- or toxicant-treated cells (with LIF present) are further normalized to that of transgenic reporter ESCs without drugs (LIF+ control). To evaluate the cell growth rate, the blue fluorescence from each dose treated, 3-day cultured cells is compared to that of the time zero cells. This comparison reveals whether or not there is a dose-dependent cell growth or negative growth at higher doses of some com. For example, if the doubling time of the ESC line is 24 h under normal culture condition, any longer doubling time under drug treatment is considered to be growth inhibition.

3.5 Validation of the Hoechst Staining with the Protein Concentration Determination

Total protein determination of the cell lysate from a culture well was initially used to represent the total cell number in the well. However, protein concentration determination is tedious, time-consuming, insensitive and precludes time zero measurements when there are few cells, potentially introducing more variables from multiples steps of pipetting and most importantly, not suitable for viable cell monitoring. We have proved that the protein concentration determination can be replaced by the fluorescence plate reading of Hoechst 33342 stained nuclei. Hoechst staining of the nuclei provides with an easier, faster and viable way to monitor the total cell number in a culture well (Fig. 1).

Hoechst 33342 is a cell membrane permeable nucleus dye that can be used for staining live cells . Hoechst-stained cells emit strong blue fluorescence that can be easily determined using a fluorescent plate reader. However, the perfection of the Hoechst staining method was validated with the widely used protein concentration determination.

The validation of Hoechst staining with protein concentration determination is carried on in drug-treated cells. Since sorbitol is our positive control for forcing potency loss, differentiation gain and growth decrease on stem cells cultured under conditions that maintain potency and proliferation, we used incremental

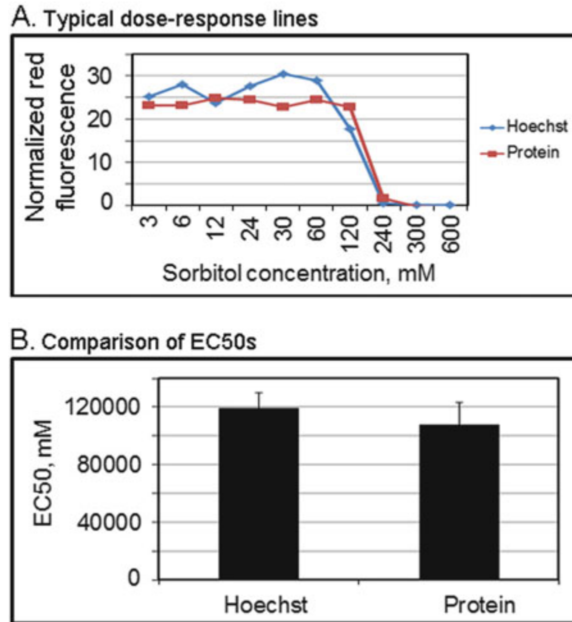

Fig. 1 Rex1 promoter reporter transgenic cells were plated and stressed on 96-well plates with incremental concentrations of sorbitol for 3 days. The cells were first recorded for blue fluorescence. Then the cells were stained with Hoechst and recorded for blue fluorescence. Panel (**a**) shows a graph of dose responses for protein and Hoechst and the EC50s are compared in panel (**b**). We use compound only, parental ESC, and Reporter ESC (Rex1-RFP ESC) 96-well plates exposed to ECVAM, and AMPK modulator toxicant sets to define compound or cellular process autofluorescence independent of reporter fluorescence, and then adjust reporter fluorescence by subtracting the two unwanted autofluorescence magnitudes

concentrations of sorbitol to do the validation. "Dose–response" graph trend lines follow logarithmic or reverse sigmoidal curves. The EC50s are calculated using Microsoft Excel software (MS Office 2010).

The reporter fluorescent readings are first normalized to the Hoechst stained nuclear fluorescence (red/blue or green/blue) or to the total protein (red/protein or green/protein). Two dose–effect lines, one normalized to Hoechst staining (currently) and the other normalized to protein concentration (previously), are obtained in this way. Correspondingly, two EC50s are calculated from the two dose–effect lines. No significant difference between the two EC50s proves the perfect validation of the Hoechst staining with the protein concentration determination. Cell autofluorescence in the Blue emission spectrum is discussed in **Note 4**.

3.5.1 Preparation of Media with Incremental Concentrations of Drugs

There are ten incremental sorbitol concentrations for positive control stress (with LIF present): 16, 23, 35, 53, 79, 119, 178, 267, 400, 600 mM, with an increment of 2/3 and a total 38-fold change in concentration. To make these concentrations, arrange ten tubes. Add 1 volume of growth medium into one to nine tube. Prepare 3 volumes of 600 mM sorbitol in the #10 tube. Transfer 2 volumes of medium from #10 tube into #9 tube, mix thoroughly. Then do the same serial dilution for the rest of the tubes.

For ECVAM drugs from a previously validation using non-transgenic mouse ESCs [33], we set up ten incremental concentrations with an increment of 1/3. To make the serial dilutions, arrange ten tubes. Add 1 volume of growth medium into one to nine tube. Prepare 2 volumes of the highest concentration of drug medium in the #10 tube. Transfer 0.5 volumes of medium from #10 tube into #9 tube, mix thoroughly. Then do the same serial dilution for the rest of the tubes. Some drugs may need a different set of dilutions for best evaluation.

3.5.2 Cell Culture and Drug Treatment

Transgenic ESCs are passaged onto 96-well plates at 10 % confluence to acclimate and grow overnight to reach ~25 % confluence at Tzero [18]. The cells are then subjected to incremental concentrations of sorbitol treatment for 3 days.

3.5.3 Testing the ECVAM Toxicant Set Previously Use to Validate Embryonic Stem Cell Toxicity Assays

The Rex1-RFP assay was based on testing of ESC responses to hyperosmotic sorbitol using embryonic stem cells [7] after hyperosmotic stress was used to characterize placental stem cell (TSCs) responses [9, 10, 15, 17] and the responses of the embryo [10, 15, 34] that give rise to ESCs and TSCs. Rex1-RFP was validated by four assays using the hyperosmotic stress dose response of Rex1-RFP in the high throughput screen (HTS) [18]. Assays by microplate reader agreed with the Rex1 protein outcomes assayed by immunoblots and reported previously [7]. Also the lack of stress-induced decrease of Oct4-GFP after 3 day of culture in the HTS here [18] agreed with the lack of a decrease in endogenous Oct4 protein at 24 h reported previously [7].

In these studies and patent (WIPO PCT; WO 2016 025510 A1) Rex1-RFP was normalized to protein to obtain stemness/ potency per cell. As described in Section 2, we next used a training set of 17 compounds previously validated in a multi-lab ECVAM study [19, 28, 29, 33] that used parental D3 ESCs previously derived by Doetschmann and colleagues [23]. Our transgenic Rex1-RFP, Oct4-GFP transgenic and parental cells are also D3 mouse ESCs [7]. Since protein assays are not sensitive enough to assay cell number at time zero in this earlier study we used a comparison of highest and lowest protein amounts at the end of culture. Our experimental design is to study toxicity effects on living cells and we exclude high doses of any compound where

the highest dose produces 10 % or less of the average protein at the three lowest doses. Since doubling rate is approximately 24 h for D3 ESCs, then there should be $2^3 = 8$ times more cells and protein in lowest stress compared with highest doses, where the dose of toxicant may decrease or stop growth from time zero on. To simplify calculations instead of 1/8th = 12.5 % we use a 10 % cutoff for high doses of toxicants that create no growth or negative growth (e.g., creating morbidity or death). In the next section we report improved direct measurements of cell number at time zero and time final using Hoechst to replace protein assays. This produces a direct measurement of growth rates. In this section we analyze the effects of 17 toxicants previously tested in and ECVAM validation [19] of toxicological responses of ESCs.

All 17 toxicants previously tested by Genschow et al. [19] in the ECVAM ESC validation caused protein loss at high doses compared with low doses and thus these slowed growth in our tests (Table 1, column 7). Approximately 76 % (13/17) ECVAM toxicants had high doses with fewer estimated cells than at time zero—that required exclusion of some high doses. The high doses producing loss of protein compared with unstressed were evenly spread in the three ECVAM groups: 4/6, 4/6, and 5/6 in the strong, weak, and nonembryotoxic groups, respectively. It should be noted that all 17 ECVAM toxicants [19, 33] were added to ESCs with LIF present under culture conditions optimized to produced potency and proliferation. Thus higher doses of toxicants may override these conditions and force loss of proliferation and potency/stemness of the ESCs (e.g., we define that "forced" means that potency loss occurs despite the presence of LIF which should maintain potency/stemness). If this occurred in an embryo one pathological response would be runting—decreased stem cell population—and a proportional increase of stem cells losing potency to compensate for fewer cells. Loss of stemness/potency indicated by Rex1-RFP decrease would indicate ability to differentiate, but actual differentiation requires a second assay and we are currently validating is a Pdgfra-GFP ESC line to test for the magnitude and subpopulation size of overt stress-forced differentiation [35]. Suffice to say that two pathologies operate due to stress, depletion of stem cell population growth by lessened doubling time and by forced differentiation. This may lead to miscarriage, runting, or diseases of later prenatal or postnatal development.

In Table 1, *Column 1* shows that 5/6 strong ECVAM embryotoxic toxicants and 4/6 weak embryotoxic toxicants caused potency loss (e.g., lowest dose subtracted from highest dose)/cell after a three day exposure. Only 1/5 nonembryotoxic compounds cause potency loss. In *column 2* an analysis of decile rankings of potency loss show that weak embryotoxic has lowest potency loss.

Table 1
Rex1-RFP HTS in ESC demonstrates differences between ECVAM strong weak and nonembryotoxic toxicants and similarities between weak and strong embryotoxic stressors

	1 Potency left at highest dose	2 Potency Loss Severity (1 = least) deciles	3 Potency loss greater than forced diff.**	4 Potency loss greater than normal diff.#	5 Potency loss IC50 μM	6 Genschow IC50 μM	7 Protein loss in potency assay IC50 μM
Strong embryotoxic							
Hydroxyurea	78 %	3	YES	YES	72	67.7	636.6
6-aminonicotinamide	93 %	1	NO	YES	4.3	23.6	50.6
5-bromo-2-deoxyuridine	1		NO	NO	21.6	6.74	105
Methotrexate	69 %	4	YES	YES	0.068	0.120	0.471
Methylmercury chloride	86 %	2	NO	YES	0.042	0.437	59
All trans retinole acid	95 %	1	NO	NO	1.4	0.991	290.3
	5 of 6		2 of 6	4 of 6			
Weak embryotoxic							
Boric acid	119 %		NO	NO	2139	4359	5190.3
Salicylic acid sodium salt	140 %		NO	NO	1184	3310	5399
Lithium chloride	78 %	3	YES	YES	10,921	14,685	25636.8
Dimethadione	83 %	2	NO	YES	204	5078	2568.9
Valproic acid	96 %	1	NO	YES	572	1691	1520.8
Methoxyacetic acid	73 %	3	YES	YES	13,607	7951	9060.4
	4 of 6		2 of 6	4 of 6			

(continued)

Table 1
(continued)

	1 Potency left at highest dose	2 Potency Loss Severity (1 = least) deciles	3 Potency loss greater than forced diff.**	4 Potency loss greater than normal diff.#	5 Potency loss IC50 μM	6 Genschow IC50 μM	7 Protein loss in potency assay IC50 μM
Nonembryotoxic							
Acrylamide	154 %		NO	NO	70.7	980	5249
Dimethyl phthalate	132 %		NO	NO	2176	3259	2008.3
D-(+)-camphor	109 %		NO	NO	22,480	4117	6083.7
Diphenhydramine hydrochloride	91 %	1	NO	NO	102	113	269.6
Penicillin-G sodium salt	149 %		NO	NO	395	2806	28389.2
	1 of 5		0 of 5	0 of 5			
					12 of 17	5 of 17	2 of 17

Column 1 shows % of potency/stemness compared with zero stress dose; *green highlights* is same or more stress than zero stress and *orange* less potency. *Column 2* is a decile ranking of potency loss. *Column 3 and 4* confirm whether toxicant + LIF has greater potency loss than maximal sorbitol positive control + LIF potency loss or LIF removal normal differentiation control potency loss, respectively. *Column 5 and 6* show IC50 for potency assay and toxicity assay by Genschow, respectively; *green highlights* show lower IC50 in potency assay and *pink* shows lower IC50 in assay by Genschow. *Column 7* shows IC50 of protein loss/cell number

** Identifies ECVAM stressors that have greater loss of potency, i.e. lower Rex1-RFP, for their highest nonmorbid dose with the highest nonmorbid dose of positive control hyperosmotic sorbitol with LIF+ differentiation control

Identifies ECVAM stressors that have greater loss of potency, i.e. lower Rex1-RFP, for the ECVAM toxicant than the LIF- normal differentiation control. This compares the highest nonmorbid dose for ECVAM toxicants

Column 3 shows that 2/6 strong and weak embryotoxic compounds cause more potency loss than maximal sorbitol mediated hyperosmotic stress—forced differentiation (e.g., with LIF present), but 0/5 nonembryotoxic compounds caused greater potency loss than stress-forced differentiation. *Column 4* shows that 4/6 strong and weak embryotoxic compounds cause more potency loss than normal differentiation (e.g., without stress but with LIF removal), but 0/5 nonembryotoxic compounds caused greater potency loss. *Columns 5 compared with 6* shows that 12/17 (71 %) compounds had a lower IC50 for potency loss than toxicity measured by Genschow et al. and 5/12 (29 %) had a higher IC50. This is expected since potency loss assays focus on submorbid events and typically occur at lesser toxicant doses.

In the Rex1-RFP assay, the nonembryotoxic group is marked by lack of potency loss compared with stress-forced differentiation, normal differentiation, and very low IC50s for change in potency compared with baseline. Strong and weak embryotoxic groups appear similar in regard to potency loss, but the strong group undergoes potency loss at low IC50 concentrations.

Toxicologically, our findings are similar to those from Stemina's metabolomic stress findings; that the ECVAM strong and weak embryotoxic groups are more similar than different [36, 37]. However, it is clear that the strong embryotoxic groups have the lowest IC50s for forced potency loss in our potency loss HTS and in the cytotoxicity assay from Genschow and colleagues. Surprisingly, since all toxicants cause cell growth rate suppression, the nonembryotoxic stresses may actually not participate in enzyme mediated compensatory differentiation described for positive control hyperosmotic sorbitol and all other stressors tested to date in TSC assays and in stress induced potency/stemness loss in two-cell stage and blastocyst stage embryos [15, 38, 39].

Taken together the data show a reduce-to-practice use of Rex1-RFP to assay stress-forced potency loss by important drug and environmental toxicants. These findings may contribute to efforts to understand developmental and reproductive toxicants for developmental biologists, reproductive scientists, pharmacologists, and toxicologists. Loss of stemness/potency due to drugs such as aspirin and metformin and diet supplements such as BR-DIM occurs in whole embryo culture of embryos producing ESCs and embryos leading to the stage when ESCs are produced [15, 38]. Rex1-RFP potency loss reporter ESCs should be able to emulate whole embryo culture and in vivo gestational embryonic toxicological responses.

3.5.4 Trend Line and EC50 Calculations for Protein- and Hoechst-Based Assays Cell Growth for Hyperosmotic Stress

To replace the insensitive protein measurements for cell number used in Section 3.5.3, fluorescent measurement of Hoechst is done. The potency reporter fluorescent readings are first normalized to the Hoechst stained nuclear fluorescence (Red Rex1-RFP/blue Hoechst) and then are normalized to the previously used total protein (Red/protein) using the standard positive control hyper-osmotic stress (e.g., sorbitol used previously for ESCs, TSCs and embryos [7, 9–11, 15, 17, 34]). Two dose–effect lines are obtained in this way. Correspondingly, two EC50s are calculated from the two dose–effect lines. No significant difference between the two EC50s proves the validation of the fluorescent plate reading of Hoechst stained cells with total protein determination (Fig. 1a, b).

3.5.5 Trend Line and EC50 Calculations for Protein- and Hoechst-Based Assays Cell Growth for AMPK Modulators

At the end of the culture with AMPK modulators (CoQ10 BR-DIM, Metformin, Salicylate/Aspirin in Fig. 2a), the cells were first read for both potency reporter fluorescence and Hoechst stained nuclear fluorescence as described in Section 3.5.4. The cells were then lysed with RIPA lysis buffer supplemented with proteinase inhibitors, 100 μL per well. The cell lysates were briefly sonicated. 10 μL of the lysate was mixed with 100 μL of BCS protein assay reagent in the well on a new 96-well plate and incubated at 37 °C for 30 min. Bovine serum albumin was used for making the

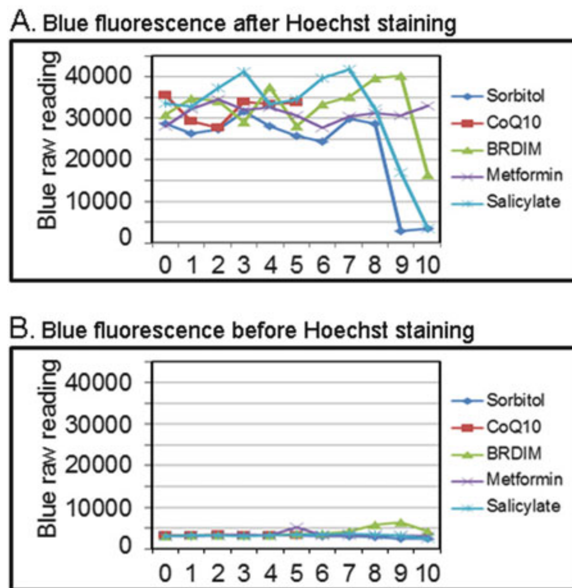

Fig. 2 Rex1 promoter reporter transgenic cells were plated and stressed on round bottom 96-well plates with incremental concentrations of drugs for 3 days. The cells were first recorded for blue fluorescence (panel **b**). Then the cells were stained with Hoechst and recorded for blue fluorescence (panel **a**)

standard at incremental concentrations of 0.08, 0.15, 0.3, 0.6, 1.25, 2.5, 5, and 10 mg/mL. Both the protein standards and the lysate samples are measured for absorbance using the plate reader. A linear equation is derived from the regression of the line using Microsoft Excel. The total protein of the lysate is calculated from the absorbance by using the linear equation.

Hoechst-stained nuclei emit strong blue fluorescent light with a raw reading range of 10^4 per well, whereas the typical raw reading of the Rex1 promoter fluorescent reporter is in the level of 10^2 per well, which means that the Hoechst staining has high sensitivity and accuracy. Due to the high sensitivity of the Hoechst protocol, the nonspecific, autofluorescence from the cellular components other than the Hoechst stained nuclei has been assessed. As shown in Fig. 2b, the autofluorescence is less than 10 % of the raw blue fluorescence reading after the Hoechst staining. We do not anticipate that need for protein assays in the future, and other transgenic reporters may also replace Hoechst staining (*see* Section 4).

3.5.6 False Hit Rate

An ideal gene expression reporter assay should report the cellular signals of interest of the study with some account for interference with signal generation or added signals from non-reporter sources. For example for interference with signal generation, he reporter should not directly bind to the treating compound and alter the reporter activity. The reporter activity also should not be affected by the normal cellular components or by the treating compound mediated irrelevant or unknown cellular components. Some reporters, unfortunately, do have high direct binding to test compounds, resulting in the quench of its activity. For example, Renilla luciferase reporter has been reported to have 40–70 % "false hit rates" to a group of 42,460 PubChem compounds [27].

In our HTS models using fluorescent proteins as the reporters, we find that there are two main sources of the autofluorescence: one is from the normal cellular components without compound treatment, the other is from the treating compound resulted in autofluorescence which needs to identified and adjusted for in a dose dependent manner for each compound (*see* **Notes 3–7, 9**).

In the stemness/potency "loss-of-function" model of HTS1, we define the fluorescence determined from the zero dose treated, normal cultured Rex1-RFP transgenic ESCs as the standard maximal reporter fluorescence. In the differentiation "gain-of-function" model of HTS2, we define the fluorescence determined from the 1 μM retinoic acid treated PDGFRα-GFP transgenic ESCs (positive control) as the standard maximal reporter fluorescence [35]. All the autofluorescence measurements determined were normalized to the standard maximal reporter fluorescence to compare the relative levels of the autofluorescence.

The false hit rate of the red or green fluorescent reporter protein was determined from the parental ESCs. The parental ESCs provide a clean system that is free of red or fluorescent protein. The autofluorescence in the absence or presence of incremental concentrations of the treating compound in parental non-transgenic cells was normalized as a percentage to the standard maximal reporter fluorescence in the transgenic reporter cells. Here we define 50 % as the threshold for false hits. For a specific treating compound, a percentage of autofluorescence that is more than 50 % is considered to be a false hit.

The false hit rate of the blue fluorescence can be determined from either the parental ESCs or from the transgenic ESCs. The blue autofluorescence was determined before Hoechst staining and the total blue fluorescence was determined after Hoechst staining. In this way, the percentage of the blue autofluorescence can be calculated from each compound at each incremental concentration point. As shown in the Table 2, both cell the parental and the transgenic cell lines exhibited similar false hit rates to the same group of 22 test compounds. These 22 compounds are 17 from the ECVAM study [19, 33] and 5 AMPK modulating drugs (Fig. 2a) and diet supplements.

Table 2
False hit rate. Transgenic or parental ESCs were cultured in 96-well plates. The cells were treated with 22 ECVAM and 5 AMPK modulator compounds (Fig. 2a) at incremental concentrations for 3 days

		0 dose	25 % dose	50 % dose
Red	False hit	0	0	0
	Total hit	22	22	22
	False hit rate (%)	0	0	0
Green	False hit	0	0	9
	Total hit	22	22	22
	False hit rate (%)	0	0	40.9
Blue	False hit	0	0	2
	Total hit	22	22	22
	False hit rate (%)	0	0	9.1
Blue	False hit	0	0	3
	Total hit	22	22	22
	False hit rate (%)	0	0	13.6

Red and green autofluorescence were determined from the parental cells. Blue autofluorescence was either determined from the parental cells (blue) or from the transgenic cells before Hoechst staining (blue*). $n = 3$. These are false hits at a 50 % threshold of highest total transgenic reporter intensity shown here for 0 dose, 25 and 50 % of the maximum dose

Autofluorescence is compound dose-dependent; at the low doses, the autofluorescence is produce by the normal cellular processes at low levels. At the highest doses, many compounds make the cells morbid or apoptotic, producing very high level of autofluorescence. Many compounds creating high autofluorescence in parental non-transgenic ESCs, have no autofluorescence on plastic only 96 well tests (for more details, *see* **Notes 3–7, 9**). For a specific compound, the 25 and 50 % responses are 25 and 50 % of the maximal response for that test compound. Here we designate three response points to define the false hit rate: (1) Zero dose response point; (2) 25 % of highest dose; (3) 50 % of highest dose. We think that the false hit rates at the 50 % of maximal response reflect the normal working concentration in this study and are relevant to the concentrations used in previous ECVAM compound studies [40].

Discussion of results from Table 2. As shown in the false hit table, the false hit rate of the red fluorescent protein to 22 compounds are 0.0 % at zero dose, 0.0 % at 25 % maximum dose, which indicates that the red fluorescent protein reporter in our Rex1-RFP transgenic ESCs is a good reporter to monitor ESC stemness and potency in our HTS1 model. The false hit rate of the green fluorescent protein to 22 compounds is 0.0 % at zero dose and 25 % of maximum doses and 40.9 % at 50 % of max dose. The relatively higher hit rate for Pdgfra-GFP HTS2 is largely due to the weakness of the green fluorescence in PDGFRα-GFP transgenic cells. By replacing with a stronger version of fluorescent protein, e.g., E2-Crimson or td-Tomato, the reporter signal will be boosted by ten times or higher. Then the increased signal–noise ratio will greatly reduce the false hit rate for the HTS2 model. Another alternative is to use a naturally low hit rate reporter to replace the fluorescent proteins in our study. For example, β-lactamase has been proved to have extremely low false hit rate [27]. These choices are discussed in Notes (Section 4) and are most important for toxicologists contemplating screening very large numbers of test compounds, many without knowledge of NOAEL, LOAEL, EC50, IC50, or LD50 dose ranges. For some other studies in pharmacology or developmental biology where the number of test compounds is low, choice of reporter may focus on whether it can be used repeatedly in viable cells without affecting normal cellular functions.

False hit: correctable or not? If a false hit is unavoidable, it should be predictable and quantitatively deductible. In the case of an enzyme as the reporter, it is difficult to quantify and correct the reporter response from the reporter inhibition by the treating compound. Switching to another reporter enzyme is usually the best way to avoid high false hit rate. For example, β-lactamase has been reported to have very low false hit rate (<0.1 %) and is more ideal than luciferase in this regard and that its secreted form has less

chance of being affected by intracellular test compound-dependent processes [27]. In our HTS model, the red and green autofluorescence can be easily measured from the parental ESCs. Therefore the autofluorescence can be deducted and corrected from the transgenic ESCs (*see* Section 4). The blue autofluorescence can be corrected either from the measurement of the blue fluorescence from the parental cells or, more directly, from the transgenic cells.

4 Notes

1. *Maintenance of the potency transgenic cell lines.* The Rex1-RFP cells, like all ESCs cultured with potency- and proliferation-maintaining LIF, are a heterogeneous population of cells in equilibrium with regard to stemness/differentiation balance and cell cycle commitment and stage. Previously Toyooka and colleagues showed that FACS-sorted Rex1−/Oct4− double reporter cells or Rex1+/Oct4+ double reporter ESCs would reestablish a population of mostly Rex1+/Oct4+ cells on culture with LIF after the isolation [41]. Although the established, irreversibly differentiated Rex1−/Oct4− become extra-embryonic endoderm and cannot dedifferentiate back into ESCs, these double negative FACS sorted cells rapidly reestablish a mostly double positive population in culture with the pluripotency maintenance growth factor LIF. We reported that ~9 % of our Rex1-RFP ESCs were parentally dim; thus had no more RFP fluorescence than non-transgenic parents. Presumably these are like the Rex1 negative cells described by Toyooka and colleagues, part of a transient shifting population of cells in equilibrium. However, we suggest that baseline zero stress dose be monitored and if absolute Rex1-RFP fluorescence magnitude decreases appreciably the transgenic ESCs be reisolated for RFP brightness using FACs. After 3 days of high but nonmorbid sorbitol doses, as much as 28 % of Rex1-RFP ESCs become as dim as parental ESCs [18]. Tests for the percentage of Pdgfra-GFP cells, and the % of parental dim Rex1-RFPs that are irreversibly committed are underway.
2. *Optimizing duration of drug or compound exposure to transgenic embryonic stem cells in the HTS*: This HTS1 model detects the change of the fluorescent reporter protein level for potency that is largely decreased by stress treatment. We have observed that the fluorescence decreases in correlation with the increase of the time and dose of stress treatment. It is critical to choose the best time point in which the extent of the decrease of potency is substantial while the cells are not kept post-confluent. The cellular level of the fluorescent reporter protein is maintained by the dynamic balance of the synthesis and the degradation of

the fluorescent protein. The Rex1 reporter protein level has been maintained in a substantially high level in ESCs under normal culture condition. Stress-mediated suppression of Rex1 promoter activity can only reduce the rate of the *new* synthesis of the fluorescent protein. In this case, the cellular degradation rate of the fluorescent protein will play a crucial role for the dynamics of its reduction. Although the specific half-life of mApple in mouse D3 ESC is not clear, the average half-life of wild-type fluorescent proteins in mouse cells has been reported to be ~26 h [42]. To get the optimal detectible signal, the total time of drug treatment must be optimized by a preliminary experiment. In our time course study using sorbitol as the positive control stressor, we have shown an accumulative loss of the RFP each day with a large total RFP loss by day 3. Since the cells became post-confluent at 3 days, day 3 was confirmed as the optimal day for dose testing [18].

As mentioned in the Methods part, a time zero plate is always set aside to evaluate the change of growth rate under stress treatment. The time zero plate enumerates the starting Hoechst-stained nuclear signal which reflects the starting confluence. We have found that the stress-induced potency loss and differentiation gain are often coincided with cellular growth inhibition due to the decrease stem cells growth being compensated by increased differentiation that requires potency/stemness loss [12, 13].

3. *Assessment of autofluorescence using ECVAM and AMPK modulating compounds.*

Autofluorescence is unwanted; nonspecific background fluorescence that adds to the desired fluorescence from the reporter or Hoechst DNA stain. Autofluorescence may come from many different sources: (1) The normal cellular molecules or processes, (2) the abnormal cellular components or processes due to drug treatment (particularly at higher doses), (3) the drug imported by cells and its metabolized products inside cells, (4) fluorescent cross-talk from one color channel to another due to the excitation/emission wavelength overlap, (5) tissue culture plastics and solutions, and (6) interactions of compounds with reporters that alter their fluorescence. Of these, category 1 is not a problem with stem cells we use, and plastics and solutions are not a problem, but should be checked if new formulations are under consideration.

Most strong and weak toxicants at higher doses increase differentiation as measured by HTS2 Pdgfra-GFP, decrease potency/stemness as measured by HTS1 Rex1-RFP and decrease cell growth as measured by Hoechst at time zero and final (Fig. 3 bottom inset). To calculate fluorescence from the reporter fluorescent proteins in HTS1 and HTS2, chemical

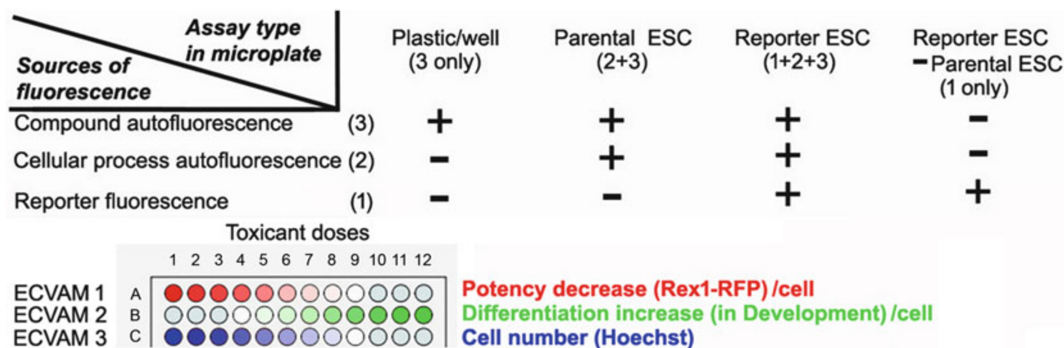

Fig. 3 We use compound-only, parental ESC, and Reporter ESC (e.g., Rex1-RFP ESC) 96-well plates exposed to ECVAM, and AMPK modulator toxicant sets to define compound autofluorescence (plastic only = 3) or cellular process autofluorescence (compound on parental cell = 2 + 3) independent of reporter fluorescence (e.g., which is 2 + 3 and 1 = reporter-dependent fluorescence), and then adjust reporter fluorescence by subtracting the two unwanted autofluorescence magnitudes. *See* Tables 3 and 4. The *inset at the bottom* shows the top three rows of a hypothetical 96-well plate with one of the set of 17 ECVAM toxicants ("ECVAM1"). Decrease in HTS1 Rex1-RFP ESCs caused by toxicant stress is detected in the top row. Increase in first lineage promoter HTS2 Pdgfra-GFP ESC is detected in the middle row and decrease in cell number by Hoechst is detected in the third row (i.e., for either HTS1or HTS2)

autofluorescence in plastic only wells and in parental ESC without reporter must be measured (Fig. 3 top). To calculate cellular autofluorescence non-transgenic, parental ESC are tested for toxicant induced effects. Both HTS1 and HTS2 fluorescent signals are normalized to time final Hoechst, and parental autofluorescence from compound or cellular autofluorescence is subtracted. Tables 3, 4, and 5 show the compound and cellular autofluorescence of 17 ECVAM compounds [19, 33] and 5 AMPK modulators (Fig. 2a) normalized to cell number by Hoechst and subtracted from HTS1 and HTS2 ESCs reporters that have both reporter and autofluorescence signals.

The relative level of autofluorescence strongly depends on the excitation and emission wavelengths used. The raw blue autofluorescence is more intense than the green; the raw green autofluorescence is more intense than the red (Table 3 panel a). Next we discuss the incidence and the magnitude of the autofluorescence in ESCs treated with 22 ECVAM toxicants [19, 33] and 5 AMPK modulating compounds in incremental concentrations.

4. Determination and Subtraction of the Blue Autofluorescence.

The blue fluorescence before Hoechst staining reflects the blue autofluorescence from the cellular molecules and processes and the drugs taken up by cells. Subtraction of the blue autofluorescence from the total blue fluorescence after Hoechst staining equals to the net blue fluorescence of Hoechst stained nuclei.

Table 3

Assessment of blue, red and green autofluorescence levels in parental ESCs (panel A), Rex1-RFP transgenic ESCs (panel B), and Pdgfra-GFP ESCs (panel C). Cells were cultured in 96-well plates with incremental doses of the 22 compounds for 3 days. Cells were read for blue, red, and green fluorescence. Then the cells were stained with Hoechst 33342 and read for blue fluorescence. Shown in this table are raw readings at the zero dose and at the highest dose without normalization. The red autofluorescence in panel B and the green autofluorescence in panel C are copied from panel A

		Parental ESCs											
		Blue autofluo* at 0 dose	Blue total at 0 dose	% of blue autofluo over blue total at 0 dose	Blue autofluo at highest dose	Blue total at highest dose	% of blue autofluo over blue total at highest dose	Red autofluo at 0 dose	Red autofluo at highest dose	Green autofluo at 0 dose	Green autofluo at highest dose		
All 22	Average value	2976	29,163	10.2	2997	16,372	18.3	14.1	16.7	503	650		
All 22	SEM value	59	713		141	3173		0.6	2.6	5.4	146		
		Transgenic Rex1-RFP ESCs											
		Blue autofluo at 0 dose	Blue total at 0 dose	% of blue autofluo over blue total at 0 dose	Blue autofluo at highest dose	Blue total at highest dose	% of blue autofluo over blue total at highest dose	Red autofluo at 0 dose (copy from parental)	Red total at 0 dose	% of red autofluo over red total at highest dose	Red autofluo at highest dose (copy from parental	Red total at highest dose	% of red autofluo over red total at highest dose
All 22	Average value	3168	26,894	11.8	2970	11,048	26.9	14.1	160.9	-8.8	16.7	62.3	-26.8
All 22	SEM value	38	732		86	1800			1.7			14.3	
		Transgenic pdgfra-GFP ESCs											
		Blue autofluo at 0 dose	Blue total at 0 dose	% of blue autofluo over blue total at 0 dose	Blue autofluo at highest dose	Blue total at highest dose	% of blue autofluo over blue total at highest dose	Green autofluo at 0 dose (copy from parental)	Green total at 0 dose	%of green autofluo over red total at highest dose	Green autofluo at highest dose (copy from parental	Green total at highest dose	% of green autofluo over red total at highest dose
All 22	Average value							503	180		675	174	
All 22	SEM value								36.2			64.1	

*autofluo is short for autofluorescence

For the blue channel, since the bandwidth of the fluorescent reporter proteins (red or green) are well separated from the blue channel, the blue autofluorescence can be assayed from the transgenic cells before Hoechst staining. The advantage of using the transgenic cells to assess autofluorescence is that the contribution of Hoechst stained nuclei is much higher than the contribution from other cellular components. As shown in Table 3 panel b, the average value of Hoechst stained nuclei in the zero dose group (baselines) is 26,894 ± 732, in which the blue autofluorescence contribution is 3168 ± 38, about 11.8 % of the total. In the highest drug concentration end, the contribution of autofluorescence to the total is 26.9 %, higher than that of the baselines. This increase of contribution may be due to several reasons: the decrease of total cell number under high concentrations of drugs, metabolized drugs, and drug-caused cellular fluorescent changes, etc. The blue autofluorescence can be easily subtracted (corrected) from the total blue fluorescence after Hoechst staining.

In both the parental ESCs and the transgenic ESCs, the blue fluorescence falls into the same range (2976 ± 59 and 3168 ± 38 respectively at the 0 dose; 2997 ± 141 and 2970 ± 86 respectively at the highest dose (Table 3 panels a and b). This consistency of the parental and transgenic cells provides further confidence to above method of data analysis. Between the parental and Rex1-RFP transgenic cells, although the blue autofluorescence contribution at the 0 dose are at about the same level (10.2 % and 11.8 %), the increase of the blue autofluorescence contribution at the highest dose in transgenic cells (26.9 %), comparing with the parental cells (18.3 %), may be due to the contribution of the fluorescent reporter protein to the blue channel (Table 3 panels a and b).

ESCs are plated on 96-well plates and cultured to ~20 % confluent at time zero [18]. Cells are treated with incremental concentrations of drugs (here we show the 22 compounds) for 3 days. Transgenic reporter HTS1 and HTS2 ESCs are first read for red, green and blue fluorescence before Hoechst staining. Then the cells are stained with Hoechst and read for blue fluorescence again.

5. Determination and subtraction of the red or green autofluorescence.

The red or green autofluorescence should be determined from the parental ESCs. The main difference between the transgenic ESCs and the parental ESCs is the expression of the fluorescent reporter protein. Therefore, the parental ESCs can be used to assess the autofluorescence that contribute to the total red or total green fluorescence in the transgenic ESCs. The non-normalized raw reading of the red autofluorescence from the

parental ESCs range from 14.1 ± 0.6 at the 0 dose and 16.7 ± 2.6 at the highest dose, which contribute 8.8 % and 26.8 % to the total red fluorescence in the Rex1-RFP transgenic cells respectively (Table 3 panel b). Although there is a small amount of increase (from 14.1 ± 0.6 to 16.7 ± 2.6) of the autofluorescence in the trend lines, the total fluorescence trend lines still show significant decrease (from 160.9 ± 1.7 to 62.3 ± 14.3), indicating the dramatic potency loss under stress treatment. The autofluorescence is subtracted from the total fluorescence. However, it has to be noted that the red autofluorescence should be added to but not be subtracted from the total red fluorescence of the transgenic cells in the data analysis.

Similarly, the green autofluorescence from the parental ESCs ranges from 503 ± 5.4 at the 0 dose and 650 ± 146 at the highest dose. Since we have not tested the parental J1 cells that was used for making the Pdgfra-GFP transgenic ESC line, the present result of the green autofluorescence shown in panel 4 may not reflect its true level in Pdgfra-GFP transgenic ESCs (Table 3 panel c).

Parental ESCs are cultured in 96-well plates with incremental concentrations of drugs for 3 days. Cells are first read for red, green and blue autofluorescence. Then the cells were stained with Hoechst 33342 and read for the total blue fluorescence. The red or green autofluorescence determined in parental cells are normalized to the blue fluorescence after Hoechst staining. The normalized red or green fluorescence will be used for the subtraction from the normalized red or green total fluorescence obtained from the transgenic cells.

6. Normalizing net red or green fluorescence to nuclear fluorescence after subtracting parental from transgenic reporter ESCs.

The blue autofluorescence can be assessed directly from the transgenic cells. Since both the blue autofluorescence and the total blue fluorescence after Hoechst staining are determined from the same number of cells in the same well, no further normalization is necessary. The subtraction of the blue autofluorescence of the transgenic cells before Hoechst staining from the total blue fluorescence after Hoechst staining is the net blue fluorescence resulting from the total cells in that well.

Since the transgenic cell line expresses fluorescent reporter protein, it is not possible to distinguish the autofluorescence that has the same excitation/emission wavelengths as the fluorescent reporter. This type of autofluorescence should be assessed using the parental cell line from which the transgenic cell line was developed. Since the autofluorescence is determined from the parental cells in an area defined by one excitation beam and the total fluorescence is determined from the transgenic cells using another excitation beam, to compare the results from two

cell lines requires that the red or green reporter fluorescence to be normalized to the cell number that is represented by the blue fluorescence of Hoechst stained cells (Table 4).

7. Stimulation indices determined from net normalized fluorescence magnitude.

   Every drug has dose-dependent trend lines. Depending on the specific observation channel of the color, the trend line may be ascending, descending or horizontal. Stimulation factor (SI) is used to describe the maximal change of the trend line between the lowest dose and the highest dose. Here we use SI $= 1.5$ as a fudge factor to group the three types of trend lines: (1) SI > 1.5, (2) $0.67 <$ SI < 1.5, and (3) SI < 0.67. A large SI tends to introduce large error to the total fluorescence determination. By examining the groups sizes of the drugs tested, we can get an understanding of the frequencies of the drugs that affect fluorescent determination (Table 4).

   However, a drug that has an autofluorescence with SI > 1.5 does not necessarily mean a large contribution to the total fluorescence. Therefore, the relative level of the autofluorescence value compared with the total fluorescence value is used to assess the true influence of the autofluorescence. For example, 4.5 % of the 22 compounds tested in Rex1-RFP transgenic ESCs have blue autofluorescence value with SI > 1.5, but the level of their blue autofluorescence value compared with the total blue fluorescence value in the zero dose end is 11.8 % and that is not significant different higher from the average blue autofluorescence value of all the 22 compounds (11.8 %, Table 4).

   At the highest doses, especially when there is growth inhibition, the relative level of autofluorescence value over against the total fluorescence value may increase dramatically. For example, the average level of blue autofluorescence in Rex1-RFP transgenic ESCs treated with the 22 compounds over against the total blue fluorescence value is 26.9 % at the highest dose end (Table 4). In data analysis, the autofluorescence value should be subtracted from the total fluorescence value, whether it is in low level or high level.

8. Emphasizing IC50 in the robust nonmorbid dose range of text toxicants.

   One of the most important parameters in the classic toxicology to define a compound is half maximal effective concentration (IC50). Drug dose that is close or higher than IC50 inhibits the proliferation. However, an IC50 does not tell us at which point of the dose that starts to kill the cells (negative growth or fewer cells than time zero). Academic research emphasizes the study of nonmorbid doses, but commercial toxicological doses that cause death are also important to define. Our research goal is to study

Table 4
Comparison of the levels of the autofluorescence with the total fluorescence. Cells were cultured in 96-well plates with incremental concentrations of the 22 compounds for 3 days. Cells were read for blue autofluorescence and total red fluorescence. Then the cells were stained with Hoechst 33342 and read for total blue fluorescence. Shown in this table are blue raw fluorescence and normalized red fluorescence at the zero dose and at the highest dose. The red autofluorescence are a result from parental ESCs as described in the Section 3

	Blue autofluorescence[a,b]		Blue total fluorescence[a,b]		% of blue autofluorescence in total	
	Zero dose	Highest dose	Zero dose	Highest dose	Zero dose	Highest dose
Average value ± SEM from all 22	3168 ± 38	2970 ± 86	26,894 ± 732	11,048 ± 1800	11.8	26.9
Average value ± SEM from Sl > 1.5 group	3139	6459	26,566 ± 877	7487 ± 1322	11.8	86.3
#of Sl > 1.5	1		16			
% of Sl > 1.5	4.5		72.7			
Average value ± SEM from Sl < 1.5 group						
# of SI < 1.5	0		0			
% of SI < 1.5	0		0			
Average value ± SEM from 0.67 < SI < 1.5 group	3169 ± 40	2885 ± 59	27,768 ± 1506	23,772 ± 2593	11.4	12.1
# of 0.67 < SI < 1.5 group	21/22		6/22			
% of 0.67 < SI < 1.5 group	95.5		27.3			
	Red autofluorescence[c,d]		**Red total fluorescence[b,c]**		**% of red autofluorescence in total**	
	Zero dose	**Highest dose**	**Zero dose**	**Highest dose**	**Zero dose**	**Highest dose**
Average value ± SEM from all 22	49 ± 2.9	229 ± 38.5	615 ± 16.2	503 ± 56.2	8.0	45.5
Average value ± SEM from SI > 1.5 group	50 ± 4.0	299 ± 42.8	655 ± 33.0	255 ± 47.7	7.6	117.3
# of SI > 1.5	15/22		8/22			
% of SI > 1.5	68.2		36.4			
Average value ± SEM from I < 1.5 group	35	25				

(continued)

Table 4
(continued)

	Blue autofluorescence[a,b]		Blue total fluorescence[a,b]		% of blue autofluorescence in total	
	Zero dose	Highest dose	Zero dose	Highest dose	Zero dose	Highest dose
# of SI < 1.5	1/22		0			
% of SI < 1.5	4.5		0			
Average value ± SEM from 0.67 < SI < 1.5 group	49 ± 3.1	58 ± 7.0	642 ± 23.9	299 ± 48.7	7.6	19.4
# of 0.67 < SI < 1.5 group	6/22		14/22			
% of 0.67 < SI < 1.5 group	27.3		63.6			

Note: (1). Blue autofluorescence is assessed from raw readings of the transgenic cells. Red autofluorescence is assessed from normalized readings (Red/Blue after Hoechst staining) of the parental cells. (2). The red autofluorescence is addition but not subtraction from red total fluorescence. Whereas blue and green need subtraction

[a] Raw fluorescent readings without normalization

[b] Rex1-RFP transgenic cells

[c] Normalized to blue after Hoechst staining

[d] Parental cells

the adverse environmental stress that changes the normal development and results in abnormal embryos, as well as the higher dose range that cause embryonic death and miscarriage. The present HTS1 and HTS2 models report compensatory and prioritized differentiation [12, 13] that focus on the dose range that inhibit the cell proliferation but does not kill the cells. Therefore, our model emphasizes life model but reports death. The reason is that living ESC lineage after stress may still contribute to the F1 and carry epigenetic markers that compromise postnatal health.

To get to know whether or not a drug dose kills the cells, a time zero plate of cells is assayed for all the spectra of fluorescent colors. This time zero plate of cells is parallel to that of the three-day drug treated cells. By comparing the blue fluorescence of the Hoechst stained cells under each dose with the time zero point, we will be able to know whether the cells have been undergoing proliferation, proliferation arrest or apoptosis/necrosis. The dose range under which the cells undergo proliferation or proliferation arrest without apoptosis/necrosis is considered to be the life model dose range.

Once the life model dose range is determined, EC50s will be calculated and examined to see whether they fall within the lifer model dose range or death model dose range. For potency reporter transgenic cells, there are two EC50s: one is calculated from the blue Hoechst stained fluorescence trend line that describes the proliferation, and the other is calculated from the reporter fluorescence trend line that describes the prioritized differentiation by potency loss and differentiation gain.

9. Assessment of autofluorescence from each compound:

Cell-free system provides a clean system to get to know whether or not and how much the autofluorescence is and whether or not the autofluorescence change with the increase of compound concentration. In the cell-free system, incremental concentrations of compounds in PBS are assayed using the plate reader. It should be noted that the three colors of autofluorescence measured from the cell-free system also include the contribution from the tissue culture plastic and the solution being used.

Generally the frequency of compounds that can increase autofluorescence above background ($SI > 1.5$ or $SI < 1.5$) is low (4.5 %, 22.7 %, 9.1 % for blue, red, and green respectively (Table 5). The average level of the compound autofluorescence (2476–2782 for blue, 9.6–12.9 for red, 350–561 for green) is at the same level of the autofluorescence determined from the parental cells (2976–2997 for blue, 14.1–16.7 for red, 503–650 for green, Table 3, panel a). Some compounds cause dose-dependent cellular autofluorescence (e.g., BRDIM) and some do not (e.g., resveratrol). However, it should be noted

Table 5

Autofluorescence from the compounds and plastic. The 22 ECVAM [19, 33] and 5 AMPK modulating compounds (Fig. 2a) were prepared as solutions in incremental concentrations in 96-well plates. The solutions were measured for blue, red, and green autofluorescence. Shown here are the fluorescence at the zero dose and the highest dose

		Compounds on plastic								
		Blue autofluo* at 0 dose	Blue autofluo at highest dose	% of blue increase at the highest dose[a]	Red autofluo at 0 dose	Red autofluo at highest dose	% of red increase at the highest dose[a]	Green autofluo at 0 dose	Green autofluo at highest dose	% of green increase at the highest dose[a]
All 22	Average value	2476	2782	12.4	9.6	12.9	34.4	350	561	60
All 22	SEM value	39	142		0.4	1.4		4	124	
All 22	# of SI > 1.5			1			5			2
All 22	% of #SI > 1.5			4.5			22.7			9.1
	Average value of SI > 1.5	2413	4698	94.7	9.3	19.7	111.8	335	2248	571
	SEM value of SI > 1.5	N/A	N/A		0.9	6.5		14	1064	
	Extreme values of SI > 1.5	2413	4698		7.3	40.7		325	3001	
All 22	# of SI < 0.67[b]			1						
All 22	% of #SI < 0.67[b]			4.5						

(continued)

Table 5 (continued)

		Compounds on plastic								
		Blue autofluo* at 0 dose	Blue autofluo at highest dose	% of blue increase at the highest dose[a]	Red autofluo at 0 dose	Red autofluo at highest dose	% of red increase at the highest dose[a]	Green autofluo at 0 dose	Green autofluo at highest dose	% of green increase at the highest dose[a]
	Average value of SI < 0.67[b]	2563	1461							
	SEM value of SI < 0.67[b]	N/A	N/A							
	Extreme value of SI < 0.67[b]	2563	1461							
All 22	#of 0.67 < SI < 1.5			20			17			20
All 22	% of 0.67 < SI < 1.5			90.9			77.3			90.9
	Average value of 0.67 < SI < 1.5	2475	2750	11.1	9.7	10.7	10.3	351	381	8.5
	SEM value of 0.67 < SI < 1.5	54	72		0.5	0.4		4	4	
	Extreme values of 0.67 < SI < 1.5	2298	3114		6	13.3		329	414	

*autofluo is short for autofluorescence

[a]% of increase at the highest dose shows the contribution of the compound only to the autofluorescence

[b]SI < 0.67 means trend line going down (decrease)

that these compound data are from cell-free, compound solutions. The true amount of compound imported by the cells may be much smaller than that in the cell-free compound solution, especially after repeated rinse of the cells with PBS before plate reading. Therefore, the autofluorescence data from the compound and plastic, cell-free system can only be a qualitative reference and cannot be used for quantitative deduction of autofluorescence from the transgenic cells. Instead, the autofluorescence data obtained from the parental cells can be used for quantitative calculation.

References

1. Puscheck EE, Awonuga AO, Yang Y, Jiang Z, Rappolee DA (2015) Molecular biology of the stress response in the early embryo and its stem cells. Adv Exp Med Biol 843:77–128
2. Watkins AJ, Lucas ES, Marfy-Smith S, Bates N, Kimber SJ, Fleming TP (2015) Maternal nutrition modifies trophoblast giant cell phenotype and fetal growth in mice. Reproduction 149:563–575
3. Canfield RE, O'Connor JF, Birken S, Krichevsky A, Wilcox AJ (1987) Development of an assay for a biomarker of pregnancy and early fetal loss. Environ Health Perspect 74:57–66
4. Wilcox AJ, Weinberg CR, O'Connor JF et al (1988) Incidence of early loss of pregnancy. N Engl J Med 319:189–194
5. Rappolee DA (1999) It's not just baby's babble/Babel: recent progress in understanding the language of early mammalian development: a minireview. Mol Reprod Dev 52:234–240
6. Sun C, Velazquez MA, Marfy-Smith S et al (2014) Mouse early extra-embryonic lineages activate compensatory endocytosis in response to poor maternal nutrition. Development 141:1140–1150
7. Slater JA, Zhou S, Puscheck EE, Rappolee DA (2014) Stress-induced enzyme activation primes murine embryonic stem cells to differentiate toward the first extraembryonic lineage. Stem Cells Dev 23:3049–3064
8. Xie Y, Zhou S, Jiang Z et al (2014) Hypoxic stress induces, but cannot sustain trophoblast stem cell differentiation to labyrinthine placenta due to mitochondrial insufficiency. Stem Cell Res 13:478–491
9. Awonuga AO, Zhong W, Abdallah ME et al (2011) Eomesodermin, HAND1, and CSH1 proteins are induced by cellular stress in a stress-activated protein kinase-dependent manner. Mol Reprod Dev 78:519–528
10. Zhong W, Xie Y, Abdallah M et al (2010) Cellular stress causes reversible, PRKAA1/2-, and proteasome-dependent ID2 protein loss in trophoblast stem cells. Reproduction 140:921–930
11. Liu J, Xu W, Sun T et al (2009) Hyperosmolar stress induces global mRNA responses in placental trophoblast stem cells that emulate early post-implantation differentiation. Placenta 30:66–73
12. Rappolee DA, Xie Y, Slater JA, Zhou S, Puscheck EE (2012) Toxic stress prioritizes and imbalances stem cell differentiation: implications for new biomarkers and in vitro toxicology tests. Syst Biol Reprod Med 58:33–40
13. Rappolee DA, Awonuga AO, Puscheck EE, Zhou S, Xie Y (2010) Benzopyrene and experimental stressors cause compensatory differentiation in placental trophoblast stem cells. Syst Biol Reprod Med 56:168–183
14. Yang Y, Arenas-Hernandez M, Gomez-Lopez N et al (2016) Hypoxic stress forces irreversible differentiation of a majority of mouse trophoblast stem cells to a giant cell fate despite FGF4. Biol Reprod (Accepted)
15. Xie Y, Awonuga A, Liu J, Rings E, Puscheck EE, Rappolee DA (2013) Stress induces AMPK-dependent loss of potency factors Id2 and Cdx2 in early embryos and stem cells [corrected]. Stem Cells Dev 22:1564–1575
16. Xie Y, Abdallah ME, Awonuga AO, Slater JA, Puscheck EE, Rappolee DA (2010) Benzo(a) pyrene causes PRKAA1/2-dependent ID2 loss in trophoblast stem cells. Mol Reprod Dev 77:533–539
17. Zhong W, Xie Y, Wang Y et al (2007) Use of hyperosmolar stress to measure stress-activated protein kinase activation and function in human HTR cells and mouse trophoblast stem cells. Reprod Sci 14:534–547

18. Li Q, Gomez-Lopez N, Drewlo S et al (2016) Development and validation of a Rex1-RFP potency activity reporter assay that quantifies stress-forced potency loss in mouse embryonic stem cells. Stem Cells Dev 25:320–328
19. Genschow E, Spielmann H, Scholz G et al (2002) The ECVAM international validation study on in vitro embryotoxicity tests: results of the definitive phase and evaluation of prediction models. European Centre for the Validation of Alternative Methods. Altern Lab Anim 30:151–176
20. Chandler KJ, Barrier M, Jeffay S et al (2011) Evaluation of 309 environmental chemicals using a mouse embryonic stem cell adherent cell differentiation and cytotoxicity assay. PLoS One 6, e18540
21. van Dartel DA, Pennings JL, van Schooten FJ, Piersma AH (2010) Transcriptomics-based identification of developmental toxicants through their interference with cardiomyocyte differentiation of embryonic stem cells. Toxicol Appl Pharmacol 243:420–428
22. Theunissen PT, Pennings JL, van Dartel DA, Robinson JF, Kleinjans JC, Piersma AH (2013) Complementary detection of embryotoxic properties of substances in the neural and cardiac embryonic stem cell tests. Toxicol Sci 132:118–130
23. Doetschman TC, Eistetter H, Katz M, Schmidt W, Kemler R (1985) The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. J Embryol Exp Morphol 87:27–45
24. Hosler BA, Rogers MB, Kozak CA, Gudas LJ (1993) An octamer motif contributes to the expression of the retinoic acid-regulated zinc finger gene Rex-1 (Zfp-42) in F9 teratocarcinoma cells. Mol Cell Biol 13:2919–2928
25. Okumura-Nakanishi S, Saito M, Niwa H, Ishikawa F (2005) Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J Biol Chem 280:5307–5317
26. Huang R, Xia M, Sakamuru S et al (2016) Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization. Nat Commun 7:10425
27. Ho PI, Yue K, Pandey P et al (2013) Reporter enzyme inhibitor study to aid assembly of orthogonal reporter gene assays. ACS Chem Biol 8:1009–1017
28. Scholz G, Pohl I, Genschow E, Klemm M, Spielmann H (1999) Embryotoxicity screening using embryonic stem cells in vitro: correlation to in vivo teratogenicity. Cells Tissues Organs 165:203–211
29. Spielmann H, Genschow E, Scholz G et al (2001) Preliminary results of the ECVAM validation study on three in vitro embryotoxicity tests. Altern Lab Anim 29:301–303
30. Artus J, Kang M, Cohen-Tannoudji M, Hadjantonakis AK (2013) PDGF signaling is required for primitive endoderm cell survival in the inner cell mass of the mouse blastocyst. Stem Cells 31:1932–1941
31. Artus J, Piliszek A, Hadjantonakis AK (2011) The primitive endoderm lineage of the mouse blastocyst: sequential transcription factor activation and regulation of differentiation by Sox17. Dev Biol 350:393–404
32. Artus J, Panthier JJ, Hadjantonakis AK (2010) A role for PDGF signaling in expansion of the extra-embryonic endoderm lineage of the mouse blastocyst. Development 137:3361–3372
33. Genschow E, Scholz G, Brown N et al (2000) Development of prediction models for three in vitro embryotoxicity tests in an ECVAM validation study. In Vitr Mol Toxicol 13:51–66
34. Xie Y, Zhong W, Wang Y et al (2007) Using hyperosmolar stress to measure biologic and stress-activated protein kinase responses in preimplantation embryos. Mol Hum Reprod 13:473–481
35. Li Q, Louden E, Dai J et al (2016) Stress forces first lineage differentiation of mouse ESCs, validation of a high throughput screen for toxicant stress. Development (manuscript in preparation)
36. West PR, Weir AM, Smith AM, Donley EL, Cezar GG (2010) Predicting human developmental toxicity of pharmaceuticals using human embryonic stem cells and metabolomics. Toxicol Appl Pharmacol 247:18–27
37. Kleinstreuer NC, Smith AM, West PR et al (2011) Identifying developmental toxicity pathways for a subset of ToxCast chemicals using human embryonic stem cells and metabolomics. Toxicol Appl Pharmacol 257 (1):111–21
38. Bolnick A, Abdulhasan M, Kilburn B et al (2016) Commonly used fertility drugs, a diet supplement, and stress force AMPK-dependent block of stemness and development in cultured mammalian embryos. J Assist Reprod Genet 33:1027–1039
39. Bolnick A, Abdulhasan, M, Kilburn B et al (2016) 2-cell embryos are more sensitive than blastocysts to AMPK-dependent suppression of anabolism and potency/stemness by commonly used drugs, a diet supplement and stress. J Assist Reprod Genet (in preparation)

40. Genschow E, Spielmann H, Scholz G et al (2004) Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. Altern Lab Anim 32:209–244
41. Toyooka Y, Shimosato D, Murakami K, Takahashi K, Niwa H (2008) Identification and characterization of subpopulations in undifferentiated ES cell culture. Development 135:909–918
42. Corish P, Tyler-Smith C (1999) Attenuation of green fluorescent protein half-life in mammalian cells. Protein Eng 12:1035–1040

Methods in Pharmacology and Toxicology (2017) 313–314
DOI 10.1007/978-1-4939-7208-1
© Springer Science+Business Media LLC 2017

INDEX

A

Alternative assays 125, 236, 237, 253
Alternative model 90, 236, 237, 247, 261–276
Apoptosis 25, 28, 30, 31, 34, 192, 211, 224, 263, 266, 268–270, 276, 307

B

Biomarkers 104, 189, 195–197, 221, 244

C

Caenorhabditis elegans 261–276
Cataract 27, 29, 30
Chemical(s) 4, 8, 9, 19, 23, 46, 47, 55, 60, 62, 64, 102, 103, 107, 108, 110, 117–178, 187, 189–193, 195–197, 205–226, 249, 261–265, 267–269, 275, 276, 299–300
Childhood radiation 26
Compensatory differentiation 293, 307
Congenital rubella...27, 29, 30
Critical periods 17–35, 49, 128

D

DART. *See* Developmental and reproductive toxicity (DART)
Developmental 6, 7, 11, 19–24, 28, 30–35, 43–110, 117–178, 208, 212, 217, 221, 223, 235–253, 261–276, 281, 293
Developmental and reproductive toxicity (DART) 62, 64, 71, 73, 84, 86, 87, 90–94, 99, 117–178, 237
Developmental and reproductive toxicology 43–110, 117–178, 261–276, 281
Developmental toxicity 23, 51, 52, 55, 62–110, 118, 120, 122–129, 131, 136–152, 235, 236, 239, 241, 242, 244–248, 253
Dose selection 55–57, 66, 73–75, 95, 132, 137, 139, 159, 160, 174
Drug disposition 1–4, 9, 12–13
Drug–drug interaction 1–13

E

EFD. *See* Embryo-fetal development (EFD)
Embryo(s).. 7–9, 12, 19–24, 28–30, 33, 34, 52, 83, 84, 90, 101–103, 105–110, 122, 142, 187, 190, 192, 193, 209, 236–242, 245–253, 267, 273, 275, 280, 281, 283, 289, 290, 293, 294, 307
Embryo-fetal development (EFD)...................... 43–110, 235–253
Embryonic stem cells (ESC) 103–105, 122, 236, 241, 242, 244, 279–310
  test................................103–105, 241–242, 244–245
Endocrine disrupting chemicals 196, 207
Endocrine disruptors 147–149, 151, 186, 187, 189, 190, 192, 195, 219, 226
EOGRTS. *See* Extended One-Generation Reproductive Toxicity Study (EOGRTS)
ESC. *See* Embryonic stem cells (ESC)
Extended One-Generation Reproductive Toxicity Study (EOGRTS)................................... 120, 123–125, 132, 144, 150, 157–178

F

Fertility and early embryonic development 48, 49, 52–53, 57, 92–94

G

Germline 220, 265, 266, 268, 269, 271–276, 284, 286
Gestational development 212
Gliogenesis 25

H

Heavy metals 185, 187, 189–191, 193
High throughput screen (HTS)................................249, 261, 264–265, 275, 276, 279–310
Hormones...3, 4, 10, 54, 59, 61–62, 77, 93, 94, 147–149, 160, 162, 164, 176, 186, 194–196, 205, 207, 210, 212–216, 219, 280
HTS. *See* High throughput screen (HTS)

I

Immunofluorescence 272–276, 281
Involution 205, 207, 210–212, 215–216, 224, 225
Ionizing radiation 21, 23–26, 35, 191, 194

J

Juvenile toxicity studies 98–101, 248

L

Lactation 2–8, 13, 46, 55, 66, 74–79, 82, 83, 85, 89, 131, 145, 146,

148, 150, 152–154, 158–162, 170, 171, 173, 174, 207, 208, 211, 214–216, 222, 224–226
Lifestyle ... 185–197, 208, 226

M

Male infertility ... 192
Mammary gland ... 166, 172, 205–226
Maternal observations ... 66–67, 76–78
Microcephaly ... 24, 27, 32

N

Neurogenesis ... 25, 252
Nonhuman primate (NHP) ... 48, 55, 64, 65, 86–101

P

Pesticides ... 119, 120, 122, 134, 144, 145, 158, 159, 185, 187, 188, 191–192, 220, 267
Phthalates ... 186, 187, 189, 195, 285, 292
Placental trophoblast stem cells (TSC) ... 219, 280–282, 289, 293, 294
Post-closure neural tube defects ... 25
Postnatal developmental toxicity ... 51, 71–85, 248
Postweaning assessments ... 79–81
Potency factor ... 281
Pregnancy ... 1–10, 12–13, 18, 19, 21, 26, 28, 29, 31, 32, 34, 44, 48–50, 52, 54, 57, 58, 66, 67, 71, 82, 83, 88–90, 94–98, 100, 137, 138, 151, 152, 193, 205, 207, 208, 211, 214–217, 221, 223, 225, 226, 247, 280, 281
Prenatal developmental toxicity ... 51, 62, 71–85, 127
Puberty ... 46, 72, 79, 83, 155–157, 160, 162, 165, 169, 170, 173, 174, 177, 191, 205, 207, 208, 210–214, 221

R

Radiations ... 21, 23–26, 35, 191, 194, 208, 224
REACH. *See* Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)
Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) ... 18, 124, 125, 132, 134, 140, 144, 158–160, 162, 168
Reproduction ... 46, 47, 52–54, 56–62, 72, 86, 90, 92, 121, 131, 132, 144–153, 155, 156, 173, 189, 190, 194, 236–237, 263, 265
Reproductive ... 21, 24, 26, 30, 43–110, 117–178, 185, 189, 191–193, 208, 220, 223, 237, 261–276, 281, 293
Retinoids ... 23, 30–35

T

Teratogen ... 23, 24, 31, 33–35, 88, 104, 108, 109, 217, 237, 244–245, 250
Teratology ... 17–35, 63–64
Testicular damage ... 185–197
Testing ... 45–48, 54, 55, 64, 70, 71, 80, 81, 86–94, 99–110, 117–119, 122–134, 136, 145, 148, 151, 152, 157, 158, 161, 172, 177, 178, 237, 239, 248–249, 253, 261, 276, 281, 289, 299
Therapeutic radiation ... 24, 25
Toxicology ... 21, 43–110, 117–178, 261–276, 281, 304
Transgenerational effect ... 192
TSC. *See* Placental trophoblast stem cells (TSC)

V

Vascular disruption ... 29

W

Whole embryo culture (WEC) ... 102, 105–108, 110, 122, 236–239, 293

Z

Zebrafish embryotoxicity test (ZET) ... 107–109