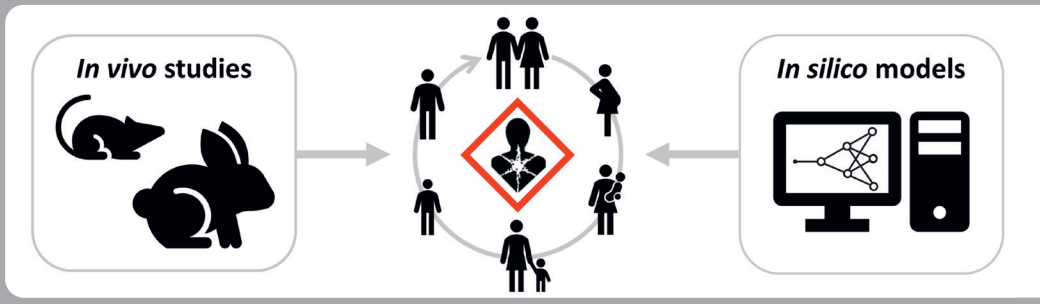
CHEMIE

Anastasia Weyrich



Challenges and solution approaches for an improved assessment of reproductive toxicity – species differences and *in silico* predictions

Dissertation
Fachrichtung Lebensmittelchemie und Toxikologie
September 2022

TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

Challenges and solution approaches for an improved assessment of reproductive toxicity

– species differences and *in silico* predictions

Vom Fachbereich Chemie der Universität Kaiserslautern zur Verleihung des akademischen Grades „Doktor der Naturwissenschaften“ genehmigte Dissertation

D 386

vorgelegt von

Anastasia Weyrich, M.Sc.

geboren in Berlin

Betreuerin: Prof. Dr. Elke Richling

Datum der wissenschaftlichen Aussprache: 13.09.2022

Promotionskommission

Vorsitzender:	Prof. Dr. Stefan Kubik
Berichterstatterin:	Prof Dr. Elke Richling
Berichterstatter:	Dr. Thomas Hofmann

Eidesstattliche Erklärung

Hiermit erkläre ich, Anastasia Weyrich, dass ich die vorliegende Arbeit mit dem Titel

„**Challenges and solution approaches for an improved assessment of reproductive toxicity – species differences and *in silico* predictions**“

selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Alle wissenschaftlich verwendeten Zitate oder Inhalte anderer Verfasser wurden ausdrücklich als solche gekennzeichnet. Kooperationen mit Mitarbeitern anderer Institutionen wurden ausdrücklich genannt. Teile dieser Arbeit wurden bereits in Vorträgen, Postern und Publikationen veröffentlicht. Die vorliegende Dissertation habe ich weder im Ganzen noch in Teilen als Prüfungsarbeit bei einem anderen Fachbereich eingereicht und kein anderes Promotionsverfahren bei einer Hochschule beantragt oder eröffnet. Die geltende Promotionsordnung des Fachbereichs Chemie der Universität Kaiserslautern vom 07. Oktober 1999 ist mir bekannt.

Kaiserslautern, den _____

Anastasia Weyrich

The practical work of this dissertation was performed at the laboratories of the departments “Experimental Toxicology and Ecology” and “Specialties and Bioscience Research” of BASF SE in Ludwigshafen, Germany.

Part of this work has already been published in peer-reviewed journals:

Weyrich, A., Frericks, M., Eichenlaub, M., Schneider, S., Hofmann, T., Van Cruchten, S., & van Ravenzwaay, B. (2022). Ontogeny of renal, hepatic, and placental expression of ATP-binding cassette and solute carrier transporters in the rat and the rabbit. *Reproductive Toxicology*, 107, 1-9. https://doi.org/10.1016/j.reprotox.2021.10.005

Weyrich, A., Joel, M., Lewin, G., Hofmann, T., & Frericks, M. (2022). Review of the state of science and evaluation of currently available *in silico* prediction models for reproductive and developmental toxicity: A case study on pesticides. *Birth Defects Research*, 1– 31. https://doi.org/https://doi.org/10.1002/bdr2.2062

Table of content

Summary

Within toxicology, reproductive toxicology is a highly relevant and socially particularly sensitive field. It encompasses all toxicological processes within the reproductive cycle and therefore includes many effects and modes of action. This makes the assessment of reproductive toxicity very challenging despite the established in vivo studies. In addition, the in vivo studies are very demanding both in terms of their conduct and interpretation, and there is scope for decision-making on both aspects. As a result, the interpretation of study results may vary from laboratory to laboratory. For the final classification, the assessment of relevance for men is decisive. The problem here is that relatively little is known about the species differences between men and the usual test animals (rat and rabbit). The rabbit in particular has hardly been researched in molecular biology. The aim of the dissertation was to develop approaches for a better assessment of reproductive toxicity, with two different foci:

The first aim was to investigate species differences, focusing on the expression of xenobiotic transporters during ontogeny. Xenobiotic transporters, of the superfamily of ATP-binding cassette transporters (ABC) or solute carriers (SLC), are known to transport exogenous substances in addition to their endogenous substrates and therefore play an important role in the absorption, distribution and excretion of xenobiotics. Species differences in kinetics can in turn have a major impact on toxic effects. In the study, the expression of 20 xenobiotic transporters during ontogeny was investigated at the mRNA level in the liver, kidney and placenta of rats and rabbits and compared with that of men. This revealed major differences in the expression of the transporters between the species. However, further studies on the functionality and activity of the xenobiotic transporters are needed to fully assess the kinetic impact of the observed species differences. Overall, the study provides a valid starting point for further systematic investigations of species differences at the protein level. Furthermore, it provides previously unavailable data on the expression of xenobiotic transporters during ontogeny in rabbits, which is an important step in the molecular biological study of this species.

The second part focused on investigating the predictive power of in silico models for reproductive toxicology in relation to pesticides. Both the commercial and the freely available models did not perform adequately in the evaluation. Three reasons could be identified for this: 1. many pesticides are outside the chemical space of the models, 2. different definition/assessment of reproductive toxicity and 3. problems in detecting similarity between molecules. To solve these problems, an extension of the databases on reproductive toxicity in relation to pesticides, respecting a uniform nomenclature, is needed. Furthermore, endpoint-specific models should be developed which, in addition to the usual structure-based fingerprints, use descriptors for, for example, biological activity.

Overall, the dissertation shows how essential it is to further research the modes of action of reproductive toxicity. This knowledge is necessary to correctly assess in vivo studies and their relevance to men, as well as to improve the predictive power of in silico models by incorporating this information.

Zusammenfassung

Innerhalb der Toxikologie ist die Reproduktionstoxikologie ein hochrelevantes und gesellschaftlich besonders sensibles Fachgebiet. Sie umfasst alle toxikologischen Vorgänge innerhalb des Fortpflanzungszyklus und beinhaltet daher eine große Zahl an Effekten und Wirkmechanismen. Dies macht die Bewertung der Reproduktionstoxizität trotz der etablierten *in vivo* Studien sehr herausfordernd. Dazu kommt, dass die *in vivo* Studien sowohl bezogen auf ihre Durchführung als auch Interpretation sehr anspruchsvoll sind und es bei beiden Aspekten Entscheidungsspielräume gibt. Dies kann dazu führen, dass die Interpretation von Studienergebnissen von Labor zu Labor variiert. Für die abschließende Einstufung ist die Bewertung der Relevanz für den Menschen entscheidend. Problematisch dabei ist, dass relativ wenig über die Speziesunterschiede zwischen Menschen und den üblichen Versuchstieren (Ratte und Kaninchen) bekannt ist. Gerade das Kaninchen ist molekularbiologisch kaum erforscht. Ziel der Dissertation war es Lösungsansätze zur besseren Bewertung der Reproduktionstoxizität zu entwickeln, wobei zwei unterschiedlichen Schwerpunkte gesetzt wurden:

Das erste Ziel war es, die Speziesunterschiede zu untersuchen, wobei der Schwerpunkt auf der Expression von xenobiotischen Transportern während der Ontogenese lag. Xenobiotische Transporter, der Superfamilie der ATP-bindenden Kassettentransporter (ABC) oder Solute Carrier (SLC), sind dafür bekannt, exogene Substanzen zusätzlich zu ihren endogenen Substraten zu transportieren und spielen daher eine wichtige Rolle bei der Absorption, Distribution und Exkretion von Xenobiotika. Speziesunterschiede in der Kinetik können wiederrum einen großen Einfluss auf die toxische Wirkung haben. In der Studie wurde die Expression von 20 xenobiotischen Transportern während der Ontogenese auf mRNA-Level in Leber, Niere und Plazenta von Ratten und Kaninchen untersucht und mit der des Menschen verglichen. Hierbei zeigten sich große Unterschiede in der Expression der Transporter zwischen den Spezies. Um die kinetischen Auswirkungen der beobachteten Artenunterschiede vollständig beurteilen zu können, sind jedoch weitere Studien zur Funktionalität und Aktivität der Fremdstofftransporter erforderlich. Insgesamt bietet die Studie einen validen Ausgangspunkt für weitere systematische Untersuchungen von Artenunterschieden auf Proteinebene. Darüber hinaus liefert sie bisher nicht verfügbare Daten zur Expression von xenobiotischen Transportern während der Ontogenese im Kaninchen, was einen wichtigen Schritt in der molekularbiologischen Untersuchung dieser Spezies darstellt.

Im zweiten Teil lag der Schwerpunkt auf der Untersuchung der Vorhersagekraft von *in silico* Modellen für Reproduktionstoxikologie in Bezug auf Pestizide. Sowohl die kommerziellen als auch die frei verfügbaren Modelle schnitten bei der Bewertung nicht ausreichend ab. Dafür konnten drei Ursachen ausgemacht werden: 1. Viele Pestizide sind außerhalb des chemischen

Raums der Modelle, 2. Unterschiedliche Definition/Beurteilung von Reproduktionstoxizität und 3. Probleme bei der Detektion von Ähnlichkeit zwischen Molekülen. Zur Lösung dieser Probleme ist eine Erweiterung der Datenbanken zur Reproduktionstoxizität in Bezug auf Pestizide, unter Beachtung einer einheitlichen Nomenklatur, nötig. Zudem sollten endpunktspezifische Modelle entwickelt werden, welche zusätzlich zu den üblichen strukturbasierten Fingerprints, Deskriptoren für zum Beispiel biologische Aktivität verwenden.

Insgesamt zeigt die Dissertation, wie essenziell es ist, die Wirkmechanismen der Reproduktionstoxizität weiter zu erforschen. Dieses Wissen ist notwendig, um *in vivo* Studien und deren Relevanz für den Menschen korrekt zu beurteilen, sowie die Vorhersagekraft von *in silico* Modellen durch Einbeziehung dieser Informationen zu verbessern.

1 Aim of dissertation

Reproductive toxicology is a highly relevant area within toxicology and at the same time one of the most challenging in terms of assessment. In this dissertation, two different aspects of the assessment of reproductive toxicity (reprotoxicity) were considered, each part being based on a publication:

The first part deals with the analysis of species differences related to the ontogeny of renal, hepatic, and placental expression of xenobiotic transporters in the rat and the rabbit. The aim was to fill data gaps (especially for the rabbit) and to compare transporter gene expression data between man, rat and rabbit using bioinformatic tools, leading to a better understanding of potential species-specific differences in developmental toxicity.

The second part deals with the prediction of reprotoxicity using *in silico* methods. Here, the reliability of known models was examined using a pesticide database, its weaknesses were identified and solution approaches for improving the predictions were worked out.

The following section introduces the scientific context on which the dissertation is based. The following three questions are addressed: “What is reproductive toxicology about?”, “How reprotoxicity is tested?” and “What factors influence reprotoxicity and how can the causes of reprotoxicity be described?”. Answering these questions illustrates the relevance of both publications as approaches to improving the assessment of reprotoxicity.

2 Introduction

2.1 What is reproductive toxicology about?

Reproductive toxicology is the study of occurrence of adverse effects on the male and female reproductive system and on development of the offspring after exposure to a substance [1]. It includes the entire reproduction cycle from formation and maturation of gametes through mating and conception, the embryonic and foetal development, postnatal adaptations, up to sexual maturation of the offspring (see figure 1) [2].

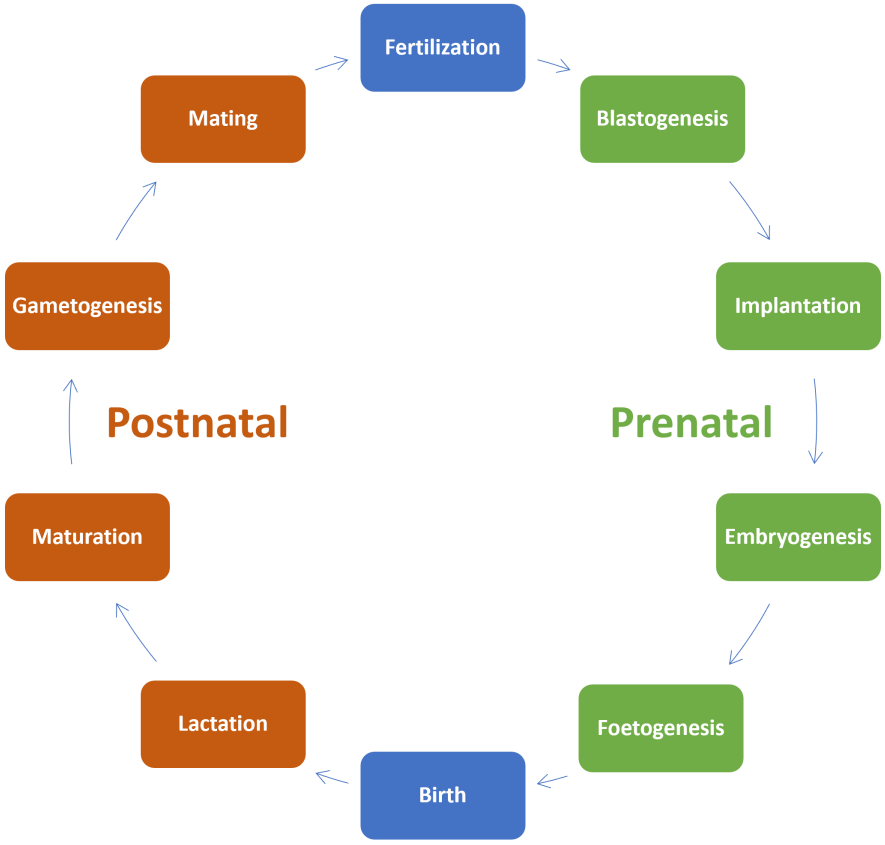


Figure 1: The reproduction cycle can be divided in two phases: the prenatal (green) and the postnatal period (orange). The prenatal period describes the development during gestation from fertilization to birth, which is the starting point of the postnatal phase. Figure was modified according to Hofmann (2013) [3].

The single steps of the human reproductive cycle are explained in more detail in the following section.

2.1.1 Reproductive and developmental biology

The reproductive cycle begins after mating with successful fertilisation. This produces the zygote, which is inaccessible to further sperm. Then the period of blastogenesis begins: after the fusion of the cell nuclei of egg and sperm, the zygote begins to divide. After about 4 days, the blastocyst develops, which already contains different cell types through differentiation. The trophoblasts initiate invasion of the endometrium, while the embryoblasts largely form the growing embryo. [4]

With the implantation of the blastocyst, embryogenesis begins. During that phase the risk of malformations due to external influences is greatest. The embryoblast initially develops into a bilaminar germinal disc consisting of the hypoblast (primitive/visceral endoderm) and the epiblast (cylindrical epithelium). During gastrulation, the bilaminar germinal disc further differentiates into a trilaminar germinal disc as cells flow in between the two pre-existing germinal discs via the primitive streak. The resulting germ layers are called the ectoderm, mesoderm, and endoderm. At the end of this period, various systems are already differentiated: large parts of the central (neural tube) and peripheral nervous systems (neural crest) are formed and precursors of the muscular system, the axial skeleton and the skin appear at the level of the somites. Subsequently, specific tissues and organs emerge from each of the three germ layers. This is the period of organogenesis in which the embryo assumes its human form. Based on morphological characteristics of the embryo, it can be divided into the Carnegie stages. This also enables the comparison of embryos between different species. There are a total of 23 Carnegie stages from fertilisation to the end of embryogenesis. [5]

The subsequent foetal period is mainly characterised by growth and differentiation of the organs that were formed during organogenesis. In men, the foetal period comprises approximately the last two trimesters of pregnancy [6]. In rats and rabbits, on the other hand, the embryonic period lasts until day 17.5 and 18.5 of gestation, for a total gestation of 21-23 and 30-32 days and, thus, accounts for more than half of gestation [7, 8].

Even after birth, development is not yet completed. In addition to growth and further development, the development of full sexual function plays a decisive role. Gametogenesis, which describes the development of the male and female germ cells (spermatogenesis [9] and oogenesis [10]), is of great importance for this. In both sexes, the maturation of the germ cells begins prenatally, but is then interrupted and does not continue until puberty.

The placenta, as well as the maternal and embryonic/foetal liver and kidney play an important role in the distribution and excretion of exogenous substances during gestation [11-14] and thus also influence the developmental toxicity of these substances. In the following, the function and development of these three organs in men are described in more detail and species differences between men and the common laboratory animals, rat, and rabbit, are discussed.

2.1.1.1 Placenta

The placenta is a temporary organ that connects the maternal with the embryonic/foetal circulation and is composed of both embryonic and maternal tissue. However, the exchange of substances is not unlimited and uncontrolled, but is regulated by the placental barrier. This can be passed through diffusion, pinocytosis, active transport, or diapedesis. The placenta in

mammals has several functions: Gas exchange, nutrition of the embryo/foetus, disposal of excretory products, barrier to harmful influences and endocrine organ (progesterone, oestrogen, gonadotrophins). Thus, the placenta takes over the absorptive function of the intestine, the secretory function of the kidney and the respiratory function of the lungs for the embryo/foetus. The placenta develops in parallel with the embryo/foetus and, therefore, changes its characteristics during pregnancy. [15]

The formation of the human placenta begins with the implantation of the blastocyst into the endometrium on day 6. The trophoblast, which is on the side of the embryoblast, invades the endometrium. The trophoblast is divided into two parts: syncytiotrophoblast (outside) and cytotrophoblast (inside). From day 9 onwards, lacunae develop in the syncytiotrophoblast, which flow together to form a labyrinth. After the syncytiotrophoblast has opened maternal vessels, the lacunae fill with maternal blood (intervillous space). Chorionic villi then form. Initially, these consist only of the two trophoblast layers (primary villi) and soon acquire an internal framework of mesenchymal tissue of the extraembryonic mesoderm (secondary villi). From the end of the 3rd week, blood vessels (tertiary villi) develop in the villi, which one week later are connected to the already functioning blood circulation of the embryo via the adhesive stalk. In the first weeks, the germinal element is surrounded all around by chorionic villi. Further villous growth is limited to the embryonic pole, while the other villi become desolate. The definitive placenta forms there by the fourth month. This is divided into the basal plate, the chorionic plate and the intervening villous trees with the intervillous space. [5]

The placentas of mammals differ among themselves in terms of their shape, type of implantation, foetal membranes and so on. The most important differences between men, rats and rabbits are shown comparatively in table 1 (more detailed information can be found in the following publications: [16-21]). The structural species differences may have an influence on the embryonic/foetal exposure of substances and thus on toxicity.

Table 1: Species differences of definitive placenta.

Species	**Chorioallantoic placenta**				**Yolk sac placenta**	**Mode of implantation**
	Gross shape	**Chorionic surface**	**Histological structure**			
			Type	**Number of trophoblast layers**		
Man	Discoid	Villous	Haemochorial	One	Becomes vestigial after the first trimester	Interstitial
Rat	Discoid	Labyrinthine	Haemochorial	Three	Inverted visceral yolk sac functions to term	Eccentric (early) Interstitial (late)
Rabbit	Discoid	Labyrinthine	Haemochorial	Two	Inverted visceral yolk sac functions to term	Eccentric (early) Interstitial (late)

2.1.1.2 Kidney

The kidney is the most important organ for the excretion of end products of metabolism and toxic substances, which are excreted through the urine. In addition, the kidney has the following tasks:

- Osmoregulation (regulation of the water balance)
- Regulation of the acid-base balance (pH value of the blood)
- Long-term regulation of blood pressure (volume regulation)
- Regulation of homeostasis (electrolyte balance)
- Production of hormones
    - Renin (short-term regulation of blood pressure)
    - Erythropoietin (stimulation of blood formation)
    - Calcitriol (active form of vitamin D)
    - Kinins
    - Prostaglandins

The kidney is paired in all mammals and located in the lumbar region to the right and left of the spine. It has the basic shape of a bean and is reddish brown in colour. The entry point of the renal artery, renal vein, lymph vessels, nerves and the exit point of the ureter is called hilus. The kidney itself consists of renal parenchyma, which is divided into the outer renal cortex and the renal medulla, which is directed inwards towards the hilum. The medulla has the shape of pyramids with their base pointing outwards and their tip (papilla) pointing inwards towards the hilum. These papillae extend freely into the cavity of the renal calices, which join together in variable form to form the renal pelvis, from which the ureter emerges. In this arrangement, urine flows out of the papilla towards the ureter. Depending on the nature of the renal papilla, the kidney is called monobranch (rat and rabbit) or multibranch (men).

The nephron forms the functional unit of the kidney. This consists of the renal corpuscle (glomerulus) and the tubule. The function of the glomerulus is to produce primary urine from the blood by ultrafiltration. The tubule begins at the urinary pole of the glomerulus and ends in a collecting duct. It can be divided into three sections: the proximal tubule, the loop of Henle and the distal tubule. In these, the initially primary urine is modified by means of two mechanisms: On the one hand, there is an active reabsorption of electrolytes, glucose, and residual proteins from the tubule into the blood and a passive reabsorption of water. On the other hand, there is an active secretion of urea, uric acid, creatinine, amino acids, and electrolytes from the blood into the tubular system. This is followed by further urine concentration up to the secondary urine.

The kidney develops from the intermediate mesoderm during organogenesis. First, the pronephros develops, which, however, is functionless. Subsequently, the mesonephros is

formed, which temporarily produces urine. Histologically, it consists of units which are similar in structure and function to nephrons of the adult kidney. Already during the regression of the mesonephros, the third kidney generation develops from the metanephrogenic blastema: metanephros. This differentiates together with the ureteric bud into the definitive kidney. The definitive nephrons and the renal stroma develop from the metanephrogenic blastema, whereas the renal pelvis, the renal calices and the collecting system develop from the ureteric bud. [5] In men, urine production of the first permanent nephrons begins between the eighth and tenth week of pregnancy, thus the kidney also takes on an excretory function for the first time, although the main organ of excretion remains the placenta [22]. By birth, all nephrons are formed, but this does not equate to full functionality. For example, the adult glomerular filtration rate is not reached until the age of two [12].

2.1.1.3 Liver

The liver is the central metabolic organ and the largest gland in the body. Its tasks include [23]:

- Biotransformation in three phases (serves to detoxify xenobiotics) [24]:
  - Phase 1: oxidation and reduction by enzymes of the cytochrome P450 system
  - Phase 2: conjugation with the help of various enzymes to increase water solubility and excretion
  - Phase 3: active transport of the transformed substances across the cell membrane of the hepatocytes into the bile
- Production of bile: synthesis of bile acids from cholesterol
- Degradation of bilirubin (degradation product of haemoglobin): glucuronidation of bilirubin and excretion via bile
- Haematopoiesis: main site of haematopoiesis during embryonic development
- Central organ of lipid metabolism: synthesis and degradation of lipids
- Formation of coagulation factors
- Regulation and storage of vitamins and trace elements
- Regulation of the glucose level in the blood

Macroscopically, the liver can be divided into 4 liver lobes and is supplied with blood via two vessels: the hepatic artery (oxygen-rich) and the portal vein (oxygen-poor, nutrient-rich). Histologically, the liver is made up of small structural elements, the hepatic lobules, which are 1-2 mm in diameter. They consist of hepatocytes arranged concentrically around a central vein, which consist of hepatocytes lined up like columns. The blood-filled liver sinusoids run between them. Enclosed by three hepatic lobules each, small islands of connective tissue, the portal triad, are found in the histological section. They contain the afferent blood vessels as well as the intrahepatic bile ducts. Together they form what is known as the Glisson triad.

Within the hepatic lobules, a gradient of oxygen, nutrients and endogenous and exogenous substances is formed, which also leads to different metabolic processes in the zones. [25]

Hepatocytes are the main parenchymatous cells that perform most metabolic functions. They are the majority of the total liver cell population. It consists of two poles: the narrow apical membrane domain forms a bile duct with one neighbouring cell at a time. The broad basolateral membrane domain borders on the space of Disse or the liver sinusoids and is responsible for the exchange of substances with the blood. Other cell populations include Copper cells (resident macrophages), Ito cells (fat and vitamin A storage), Pit cells (specialised lymphocytes), endothelial cells and the epithelial cells of the bile ducts. [26]

The liver arises from an epithelial bud of the embryonic foregut, which proliferates and differentiates into the mature organ. It is thus a derivative of the endoderm. The liver bud, consisting of hepatoblasts, grows strand-like into the mesenchyme of the septum transversum. The architecture of the liver then begins to establish: hepatic sinusoids and bile ducts form and the liver bud divides into lobes. The left umbilical vein becomes the ductus venosus and the right vena vitellina becomes the portal vein. The bipotent hepatoblasts begin to differentiate into biliary epithelial cells and hepatocytes. Hepatoblasts adjacent to portal vein differentiate into biliary epithelial cells and form a bilayer of cuboidal cells. In the ductal plate, focal dilations develop at points in the bilayer, surrounded by portal mesenchyme, and develop into intrahepatic bile ducts. Hepatoblasts that are not adjacent to the portal vein differentiate into hepatocytes and arrange into cords lined by sinusoidal epithelial cells and bile ducts. Thus, they begin to take over the functions of a mature hepatocyte. [5]

The development of biotransformation capacity is relevant and critical from a developmental toxicology perspective, as it has a major impact on the bioavailability of toxic substances in the foetus [27]. Studies have shown that the human foetus has a well-developed system of xenobiotic metabolising enzymes [28]. However, foetal, and neonatal liver functions are significantly reduced quantitatively compared to the adult stage and differ qualitatively in terms of active enzymes [13, 14, 22]. Thus, the liver undergoes a postnatal maturation process until it reaches its adult capacity.

The structural development of organs during the embryonic and foetal phases is described in detail in embryology. Research into the functional development of the human foetus is limited by the lack of accessibility and the ethical restrictions that stand in the way of research on the foetus, which is why there are gaps in knowledge here. However, this knowledge would be crucial for a better understanding of developmental toxicity, as toxicokinetics and toxicodynamics processes can be derived from it.

2.1.2 Types and consequences of reprotoxic effects

Within reproductive toxicology, toxic effects on the entire reproductive cycle are considered and therefore includes a variety of findings. Possible adverse effects on sexual function and fertility are for example alterations to the female and male reproductive system, an altered start of puberty and gamete production or changes in sexual behaviour and parturition [29]. An example for a reprotoxic substance is the nematicide 1,2-Dibromo-3-chloropropane (DBCD). A pesticide, which leads to male fertility disorder [30]. Developmental toxicity includes all adverse effects on the development of the offspring during prenatal and postnatal phase caused by parental exposure. The four major manifestations of developmental toxicity are mortality, dysmorphogenesis (structural alterations), alterations to growth, and functional alterations [29]. The most well-known developmental toxic substance is thalidomide, a sleeping pill that was marketed between 1957 and 1961 under the name Contergan. This is a teratogen and leads to severe deformities of the limbs if taken within the first three months of pregnancy [31].

2.1.2.1 Classification of reproductive toxicants

Reprotoxic substances are classified according to the Globally Harmonised System of Classification and Labelling of Chemicals (GHS). The classification is based on the estimated hazard to men (see table 2, hazard class category) [29]. Through the associated hazard statements, there is also a distinction between adverse effects on the sexual function and fertility (F/f) or on the development of the offspring (D/d).

Table 2: Classification and labelling for reproductive toxicants according to GHS [29]

Classification		Labelling		GHS hazard statement codes
GHS hazard class	**GHS hazard class category**	**GHS signal word**	**GHS hazard statement**	
Reproductive toxicity	**1A** (Known human reproductive toxicant)	**Danger**	May damage fertility (F) or the unborn child (D)	**H360** (state specific effect if known: F, D, FD, Fd, Df)
	1B (Presumed human reproductive toxicant)			
	2 (Suspect human reproductive toxicant)	**Warning**	Suspect of damaging fertility (f) or the unborn child (d)	**H361** (state specific effect if known: f, d, fd)

The reprotoxic potential of a substance, which can lead to a GHS classification, has a great influence on its potential use. This is explained in more detail in the following section.

2.1.2.2 Regulation of pesticides, biocides, and chemicals in EU

The approval of pesticides, biocides and chemicals is strictly regulated within the EU to ensure high safety standards for the population and the environment. The assessment of the hazard is based on toxicological studies in which potential hazards are identified and characterized. The table below lists the governmental documents for the regulation and data requirements for pesticides, biocides, and chemicals in EU regarding human health:

Table 3: Regulations and data requirements for the approval of pesticides, biocides, and chemicals

Type	Regulation	Data requirements
Pesticide (Active substance)	Regulation (EC) No 1107/2009 [32]	Commission Regulation (EU) No 283/2013
Biocide	Regulation (EC) No 528/2012 [33]	ECHA Guidance on the Biocidal Products Regulation, Volume 83 Human heath, Part A: Information Requirements
Chemical	Regulation (EC) No 1907/2006 (REACH) [34]	ECHA Guidance on Information Requirements and Chemical Safety Assessment, Chapter R.7a: Endpoint specific guidance

The European Food Safety Authority (EFSA) is responsible for the approval of pesticides and their active substances. Cooperation with the European Chemicals Agency (ECHA) is sought here, as this is responsible for creating the harmonized classification according to the GHS. In addition to the prenatal developmental toxicity study (OECD 414) in two species (usually rat and rabbit), a generational reprotoxicity study (OECD 416 or 443) is required for the authorisation of pesticides. The exact study requirements have to be discussed with the authorities individually for each substance. The classification of an active substance of the pesticide as toxic for reproduction (category 1A or 1B) means that it cannot be approved or only under special and very restricted conditions. Reprotoxicity is thus an exclusion criterion for the approval of pesticides in the EU. The following section describes assessment of reprotoxicity and explains the challenges that arise.

2.2 How reprotoxicity is tested?

2.2.1 *In vivo* test guidelines

Due to the complexity of the reproduction cycle and the resulting high number of possible effects and endpoints that have to be assessed, testing for reprotoxicity is usually carried out in sections with the help of several studies. In the EU, the studies for chemicals and pesticides follow the guidelines of the Organization for Economic Co-operation and Development (OECD). The guidelines of the Office of Prevention, Pesticides & Toxic Substances (OPPTS) of the US Environmental Protection Agency (EPA) are almost equivalent. For drugs, there are global guidelines that have been defined by the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH). Drug candidate reprotoxicity testing is divided into three segments: 1. Fertility and Early Embryonic Development (FEED) Study, 2. Embryo-Fetal Developmental (EFD) toxicity study and 3. Pre- and Postnatal Developmental (PPND) toxicity study [2]. These differ from the study guidelines for pesticides and chemicals among other things in the length of the exposure period, the dosage and administration routes. The following table summarizes the guidelines that are available for testing the reprotoxic potential of pesticides and chemicals. A more detailed description of the OECD guidelines can be found in the second publication (see 4.2).

Table 4: Comparable studies to assess reprotoxicity of chemicals and pesticides in EU or US

OECD (EU)	**US EPA**
Test No. 414: Prenatal Developmental Toxicity Study [35]	870.3700 - Prenatal Developmental Toxicity Study [36]
Test No. 416: Two-Generation Reproduction Toxicity [37]	870.3800 - Reproduction and Fertility Effects [38]
Test No. 421: Reproduction/Developmental Toxicity Screening Test [39]	870.3550 - Reproduction/Developmental Toxicity Screening Test [40]
Test No. 422: Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test [41]	870.3650 - Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test [42]
Test No. 426: Developmental Neurotoxicity Study [43]	870.6300 - Developmental Neurotoxicity Study [44]
Test No. 443: Extended One-Generation Reproductive Toxicity Study [45]	No comparable study available

The assessment of reprotoxicity based on the mentioned *in vivo* studies is very challenging, both in terms of conducting the studies and interpreting them. Since both aspects have a great influence on the classification of a substance regarding its reprotoxicity, they are discussed in the following two sections:

2.2.1.1 Challenges in conducting studies

The studies are associated with a high number of animals, a long study duration, high costs and high material consumption. Conducting reprotoxicity studies therefore requires highly experienced laboratory staff who must be able to manage the extensive study procedure and correctly record all the required endpoints.

The principal procedure of the respective study, as well as the required endpoints, are described in the guidelines. However, the guidelines do not go into detail about how the study is to be conducted, so that there is room for flexibility. An example of this is the examination of skeletal changes: For this, either only the bones or by a double staining, bones and cartilage can be stained, which can have an influence on the findings [46].

Another important point that can significantly influence the results of the study is the choice of dose in the different dose levels [47]. The guidelines stipulate at least three dose levels and a control group for reprotoxicity studies. According to OECD 414 guideline, the highest dose should meet the following requirements: "induce some developmental and/or maternal toxicity (clinical signs or a decrease in body weight) but not death or severe suffering" [35]. Of course, this description leaves room for interpretation. There is both the risk of underdosing, which can lead to underprediction of toxicity, and overdosing, which can lead to overly strict classification of a substance [48]. Due to the scope for interpretation and the high toxicological relevance of dose selection, this is a controversial topic in the scientific community, ultimately involving a trade-off between animal welfare and study validity [49].

2.2.1.2 Challenges in the interpretation of the studies

The evaluation of the reprotoxic findings is also very challenging and requires a great amount of specialist knowledge. It starts with the description of the findings. To avoid imprecise terms, Markis et al. (2009) published a glossary containing the "terminology of developmental abnormalities in common laboratory mammals" to describe findings in foetal and neonatal morphology [50]. Such glossaries form the basis for a uniform description and resulting evaluation of findings.

Additionally, a critical point of discussion is the classification of findings as malformations or variations. A malformation is defined as a permanent structural change that is likely to affect the survival or health of the species studied [51]. Variations are changes that occurs within the normal population studied and are not likely to adversely affect survival or health [51]. Variations are often developmental delays that have the potential to recover. Ten DevTox workshops discussed the correct description and classification of developmental toxicity since 1995. During these, the term “grey-zone anomalies” was introduced, which describes effects that cannot be clearly classified as malformations or variations [52]. According to Marx-Stoelting et al. (2021), the main reasons for the existence of the grey zone group are imprecise descriptive terms and insufficient knowledge of the postnatal consequences of the findings [53]. Since an evaluation of all findings is required within the study report, different estimations of findings from the group of grey zone anomalies are inevitable.

Another aspect that is controversially discussed in the interpretation of reprotoxicity studies is the influence of maternal toxicity [54, 55]. This describes maternal effects such as adverse clinical observations, reduced food consumption, decreased body weight, gross/histopathological lesions, and maternal death. When effects on the offspring occur with simultaneous maternal toxicity, the question arises as to whether the effects can be attributed directly to exposure with the substance or whether they are a consequence of the maternal toxicity [56]. It is hence decisive whether the maternal toxicity is provably necessary for the occurrence of the developmental toxicity. So far, the causal connection between a general decrease in maternal food consumption and body weight gain and the decreased foetal body weight and associated variations like reduced ossification has been admitted due to the biological plausibility [55, 57, 58]. Apart from that, little is known about the underlying mechanisms of maternal toxicity and related developmental toxicity.

Unlike concurrent control groups, historical control data are not required by regulation, but can be very useful and appropriate for interpreting study results [59]. They are used for quality assurance of the test system and to identify abnormal controls. In addition, they represent background variations and help to distinguish actual effects from random findings, as well as to determine their biological relevance [60]. The incorrect selection of historical control data can mask potential treatment effects if, for example, biological variation is increased within the

control data. Therefore, guidance has been published by the authorities that addresses the generation of historical control data: When historical control data are used for the authorisation of active substances in the EU, they are required to be strain-specific and to come from the laboratory that conducted the relevant study [61]. Furthermore, they should cover a period of five years, with the date of the study preferably in the middle of this period [61]. Practical guidance on the generation and use of historical control data in the context of reproductive and developmental toxicity studies was published by Mylchreest and Harris in 2013 [62].

The challenges just described in conducting and evaluating reproductive toxicology studies show their high complexity. This can lead to the interpretation of results varying from laboratory to laboratory.

2.2.2 Alternative methods for reproductive and developmental toxicity testing

2.2.2.1 In vitro assays

The development of alternatives to animal experiments in reproductive toxicology is becoming more and more relevant due to ethical aspects. The starting point for this development was the postulation of the 3 Rs by Russell and Burch in 1959 [63]. Since then, several *in vitro* and ex vivo methods for predicting developmental toxicity have been published, but none of them is accepted by the authorities as an alternative to *in vivo* studies. *In vitro* methods are primarily used in research and as a potential screening approach. The published *in vitro* methods are limited to the prediction of embryonic toxicity, the major drawback of these methods being the lack of interaction with the maternal compartment.

An overview of alternative methods can be found in the EURL ECVAM database on alternative methods to animal experimentation (DB-ALM) [64]. The most common alternative methods for predicting developmental toxicity are listed in the table below:

Table 5: Alternative in vitro assays for the assessment of developmental toxicity

Name	Experimental system	Biological endpoints	Ref.
Zebra fish embryotoxicity test (ZET)	zebra fish embryo culture	viability, developmental stage, gene expression, morphology	[65]
Frog embryo teratogenesis assay Xenopus (FETAX)	amphibian embryo culture	viability, apoptosis, embryo development	[66]
Whole embryo culture test (WEC)	post-implantation whole embryo culture (rat, mouse, hamster, rabbit)	cell differentiation/morphology, embryo growth/viability	[67]
Embryonic stem cell test (EST)	embryonic stem cells (human/animal origin)	apoptosis, cell cycle analysis/ differentiation/proliferation/viability, cellular functional parameters, DNA damage, gene expression profile, metabolite profile	[68]
Chicken Embryotoxicity Test (in ovo (CHEST) or ex ovo)	chick embryo in ovo or whole embryo culture (chicken)	apoptosis, cell migration, cellular functional parameters, DNA damage, embryo development/growth/morphology/viability, gene expression	[69]

2.2.2.2 In silico methods

In addition to *in vitro* assays, *in silico* methods are another way to assess the potential toxicity of a substance. With the entry into force of the REACH Regulation in 2007 [34], *in silico* methods gained in importance due to the enormous number of additional animal tests required, especially since they achieve a significantly higher throughput than animal studies (and *in vitro* experiments). Other advantages of *in silico* methods are: they save time, cost and substance, have higher reproducibility (using the same model) and can be constantly optimised (new properties, descriptors, chemical space) [70]. The application area of *in silico* prediction models is primarily in screening. In the regulatory context, the models are already used within the framework of the authorisation for metabolites and impurities of plant protection products. To ensure the quality of the models, the OECD has defined five principles that must be met by QSARs used for regulatory purposes: (1) defined endpoints; (2) unambiguous algorithm; (3) defined scope; (4) appropriate measures of goodness of fit, robustness and predictability; and (5) a mechanistic interpretation [71].

The basic idea of *in silico* prediction models is that biological activity is a function of the chemical properties of the substance [72]. A distinction is made between two types, which are described below.

- Quantitative Structure–Activity Relationship (QSAR) models
  QSAR models are based on the assumption that molecules with a similar chemical structure produce similar toxic effects [73]. They are statistical models built using a training data set of example molecules with known toxicity (workflow for building a QSAR model, see figure 2) [70]. The description of the chemical structure is based on descriptors. The choice of descriptor has a great influence on the results of the QSAR model, as they differ greatly in their description of the molecule and thus lead to different assessments of similarity. 1D or molecular descriptors are based on the geometrical, electronic, topological, constitutional and thermodynamic properties of the molecule [74]. The chemical structure of a molecule can also be described using fingerprints (2D descriptors) and is represented in the form of a bit vector. In substructure key-based fingerprints, each bit represents the presence or absence of a predefined substructure [75]. In contrast, topological or pathway-based fingerprints analyse all fragments of the molecule that follow a pathway up to a certain number of bonds, and then hash each of these paths to create the fingerprint [76]. Circular fingerprints are also hashed topological fingerprints, but they are not based on a path, but the area around each atom up to a certain radius [77]. Most current QSAR models are based on fingerprints. In addition to chemical properties and molecular structure, all sorts of other properties can be used, such as affinity towards a certain enzyme [76]. The concept of applicability domain is used to check the reliability of the results of a QSAR model. This makes it possible to estimate the uncertainty in the prediction of a particular molecule, depending on how similar it is to the compounds used to build the model [78].
- Structural alert (SA) or expert-rule based models
  SAs are chemical structures associated with toxic events [79]. The alerts can be based on expert knowledge (rule-based models) or generated by statistical evaluation [80]. The advantages of these models are that they are easy to interpret and make it possible to localise the structure that is crucial for toxicity. Limitations are that the methods only show the presence or absence of SAs and thus an incomplete SA set, can lead to many false negative predictions [81].

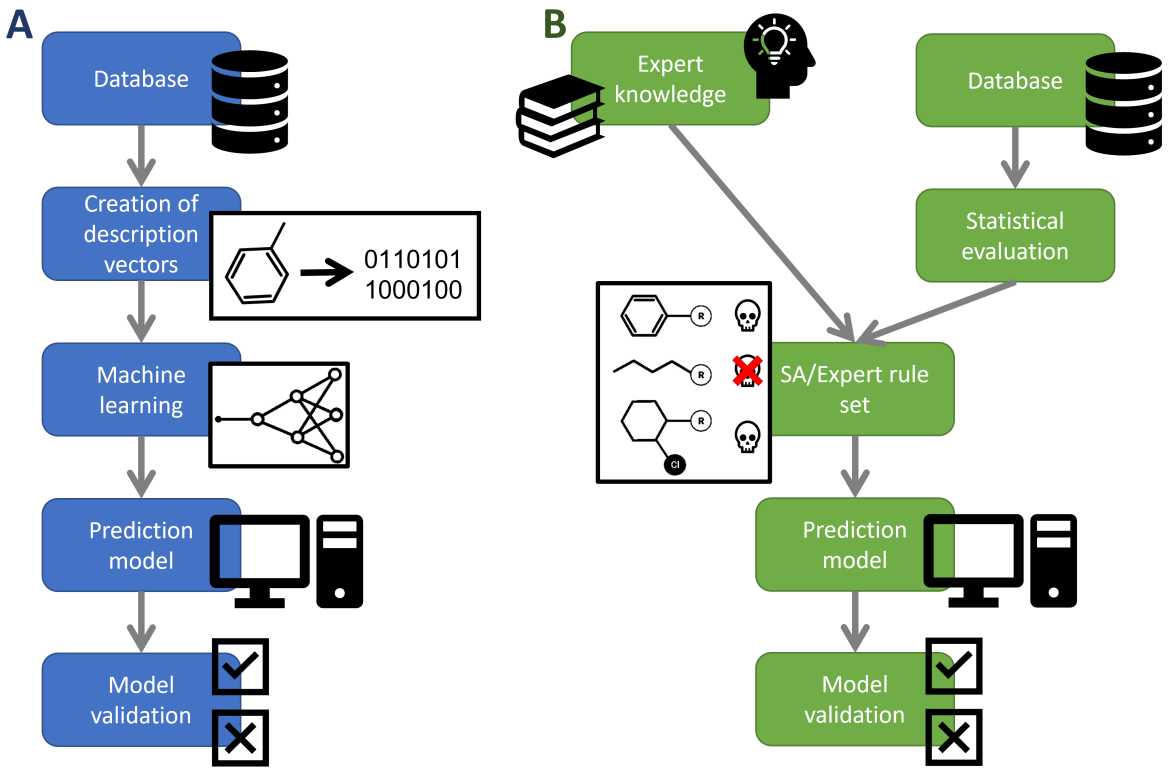


Figure 2: Workflow for building a QSAR model (A) or a SA or expert-rule based model (B).

Due to the importance of reprotoxicity in pesticide registration and the enormous animal use, time, and cost of *in vivo* studies, *in silico* prediction models for screening early research compounds for reprotoxicity are of great interest. The major challenges for *in silico* prediction of reprotoxicity endpoints are the complexity of ontogeny, the combination of multiple endpoints with partly unknown AOPs and the limited availability of empirical reprotoxicity data [82].

2.3 What factors influence reprotoxicity and how can the causes of reprotoxicity be described?

In the assessment of reprotoxicity by means of *in vivo* studies, the focus is on the recorded findings. Based on these, a hazard assessment is made in the EU. The underlying mechanisms play only a subordinate role, especially since they are often not elucidated. However, especially when assessing possible species differences between laboratory animals and men, this knowledge is of great importance, which is why research in this area is also very interesting from a regulatory point of view. The following section provides an overview of the background as well as known mechanisms, Mode of Action (MoA) and adverse outcome pathways (AOPs).

2.3.1 Toxicokinetics and toxicodynamics

For an elucidation of the causes of reprotoxicity, it is necessary to consider toxicokinetics and toxicodynamics. Toxicokinetics describes the processes to which a toxic substance is

subjected in an organism. It shows in which temporal and quantitative concentration the substance is present in different areas of an organism. This forms the basis for the toxic biochemical and physiological effects of the substance on the organism, which is described by toxicodynamics [83].

2.3.2 ADME

Pharmacokinetics is typically divided into four processes: Absorption, Distribution, Metabolism and Excretion. These are often abbreviated with the term ADME and described individually below. Oral uptake is assumed here, as this is common in reproductive toxicology studies.

In ADME, transport across membranes plays a crucial role as the substance usually passes through several cells to reach different tissues and body fluids [84]. These are often polar cells such as enterocytes, hepatocytes, or proximal tubule cells, which have an apical and a basolateral side. The apical membrane faces the lumen, in this case the intestinal lumen, bile ducts and urine, while all other cell sides are referred to as basolateral. Transport through membranes depends on the molecular structure of the substance as well as membrane-specific properties [85]. In addition to the simple diffusion processes, a distinction is made between passive and active transport, which facilitate or enable the transfer across the membrane. Passive transporters do not depend on external energy for transport, as they transport molecules only in the direction of their electrochemical gradient. These include channels and carriers. Active transporters such as symporters, antiporters and pumps, on the other hand, require external energy for transport and are therefore also able to transport molecules against their electrochemical gradient [86]. Transport across membranes is regulated by their permeability as well as the composition of various transporters and differs greatly between the apical and basolateral membranes of polar cells.

Absorption describes the uptake of the substance into the bloodstream. In the case of oral exposure, this occurs via the mucous membranes of the gastrointestinal tract [84]. This absorption process can be based on all different transport mechanisms previously described.

The substances absorbed in the gastrointestinal tract first reach the liver via the portal vein, where most substances are metabolised before they enter the systemic circulation [87]. The aim of metabolization is to improve excretion from the body and thus detoxification. In some cases, however, metabolization also results in toxicity, as in the case of methanol, for example. This itself is only slightly toxic, but metabolization produces the metabolites formaldehyde and from this formic acid, which triggers acidosis [88]. Metabolism takes place in various mucous membranes, in the intestine, in the lungs and in the blood plasma, but the main site of metabolization is the liver. Instead of entering the bloodstream after metabolization, substances can also be excreted directly from the liver via the bile into the intestine. If the substance is reabsorbed in the intestine, multiple passage through the portal system is also

possible. This shuttling of a substance between the intestine and the liver is called the enterohepatic circulation [89].

Once the absorbed substance has reached the systemic circulation, it is distributed throughout the body with the blood. When substances are distributed, their properties such as solubility, chemical structure or binding capacity to plasma and tissue proteins play an important role. Lipophilic substances, for example, tend to accumulate in fatty tissue [85]. Organ-specific properties such as their blood flow and the permeability of the surrounding membrane also have a major influence on the respective substance concentrations and retention time [87].

The excretion of substances largely takes place via the kidneys by means of urine [85]. All substances freely dissolved in plasma up to 15 kDa enter the primary urine, whereas substances bound to plasma proteins are retained in the blood. When concentrated into secondary urine, excretion depends primarily on the polarity of the substance: lipophilic substances diffuse back into the blood, while polar substances remain in the urine. In addition, there are active transporters in the proximal tubule that transport both endogenous and exogenous substances and can thus actively excrete exogenous substances into the urine or absorb them into the blood. During lactation, breast milk also plays a role for lipophilic substances because of its fat content [87]. This is relevant from a reprotoxicity point of view because exposure of the offspring to harmful substances is possible via the milk.

In the case of reproductive toxicology, in addition to maternal ADME, that of the embryo/foetus is also relevant. Here, the transport via the placenta plays a decisive role, which occurs both towards and away from the foetus. From the point of view of the embryo/foetus, the placenta serves as an organ of absorption and excretion of substances. As already described in 2.1.1, the organs are physiologically developing during ontogenesis and therefore constantly change their functional properties. Since the ADME parameters are also constantly changing, their determination is very demanding. In the publication of 2021 by Abduljalil et al. a data set that maps foetal cardiac output and tissue perfusion during development had been described for the first time and this can be used as a basis for PBPK models. Overall, too little is known about the individual ADME parameters in the foetus, so that a prediction of the exposure of the foetus to a potentially toxic substance cannot be made.

2.3.3 Description of the causes of reprotoxicity

Various concepts can be used to describe the causes of reprotoxic effects. These are often used synonymously in practice/literature, but are correctly defined as follows:

1. Mechanism of action (MOA)

   The MOA describes a specific biochemical interaction through which a substance exerts its effect. This requires a comprehensive understanding of what happens at the

molecular level [90]. MOA usually involves the naming of a specific molecular target to which the substance binds. This can be an enzyme or a receptor, for example.

2. Mode of Action (MoA)

   The MoA describes functional or anatomical changes at the cellular level that result from the exposure of a living organism to a substance. It is defined as a biologically plausible sequence of key events and processes that begins with the exposure of the organism to a substance and continues through functional and anatomical changes in biological pathways that lead to toxicological findings [90].

3. Adverse outcome pathway (AOP)

   AOPs are conceptually identical to MoA but do not apply to specific substances. Substance-specific properties such as toxicokinetics are therefore not included [91]. The difference between AOP and MoA and their structure are shown in figure 3.

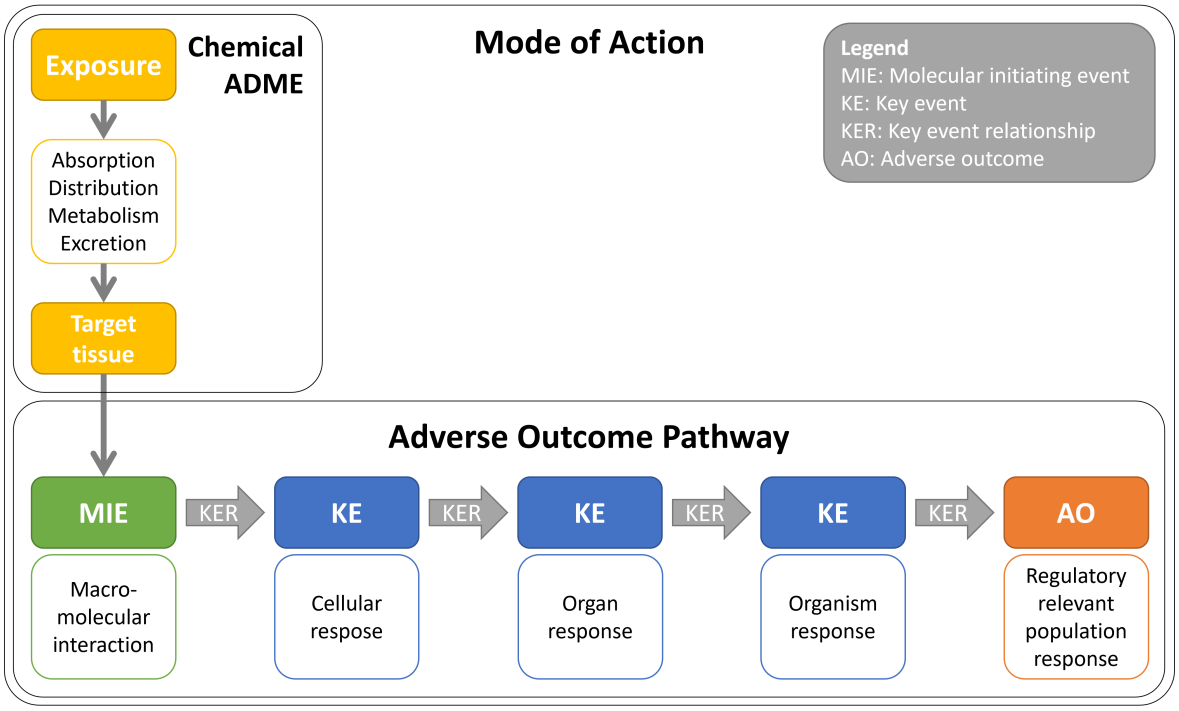


Figure 3: Relationship between MoA and AOP. An AOP consists of KEs and KERs at different levels of biological organisation, ranging from macromolecular interactions to population reactions. The MIE describes the initial interaction of the substance with the biological system. The AO describes the final adverse effect at the individual or population level. The KEs at the molecular and cellular level represent potential starting points for in vitro screening. By expanding the AOP with substance-specific information, it becomes the MoA. Modified after Edwards et al. (2016) [91]

Since 2012, the OECD has been promoting the development and use of AOPs. For this purpose, documents on the development, use and review of AOPs have been published [92-94]. AOPs at all stages of development are available in the AOP Wiki, an interactive and virtual encyclopaedia for AOP development. This also contains AOPs describing AOs that can be assigned to the area of reprotoxicity. Table 5 lists all current AOPs endorsed by the OECD

with a reprotoxic background. Considering the enormous amount of potential reprotoxic findings, only a fraction is covered by the AOPs published so far. This also reflects the large knowledge gaps that exist with regard to the causes of reprotoxicity.

Table 6: Collection of all endorsed AOPs within the AOP-Wiki with reprotoxic AO (accessed 11.06.2022).

Title	**AO**	**OECD Status**
Alkylation of DNA in male pre-meiotic germ cells leading to heritable mutations	336: Increase, Heritable mutations in offspring	WPHA/WNT Endorsed
Androgen receptor agonism leading to reproductive dysfunction (in repeat-spawning fish)	360: Decrease, Population trajectory	WPHA/WNT Endorsed
Aromatase inhibition leading to reproductive dysfunction	360: Decrease, Population trajectory	WPHA/WNT Endorsed
Aryl hydrocarbon receptor activation leading to early life stage mortality, via increased COX-2	947: Increase, Early Life Stage Mortality	WPHA/WNT Endorsed
Aryl hydrocarbon receptor activation leading to early life stage mortality, via reduced VEGF	947: Increase, Early Life Stage Mortality	WPHA/WNT Endorsed
Histone deacetylase inhibition leading to testicular atrophy	1506: Testicular atrophy	WPHA/WNT Endorsed
Inhibition of Thyroperoxidase and Subsequent Adverse Neurodevelopmental Outcomes in Mammals	402: Cognitive Function, Decreased	WPHA/WNT Endorsed

2.4 Interim conclusion

In the introduction, the basics of reproductive toxicology and its investigation were discussed in detail. Despite the established *in vivo* studies, the assessment of reprotoxicity remains challenging. The most important reasons for this are:

- Complexity of reproductive toxicology (encompasses the whole reproduction cycle, consideration of a continuously developing and thus changing system, large number of effects and MoA)
- Conduct of studies (extensive study process, high number of endpoints)
- Interpretation of studies (malformation vs. variation, maternal toxicity, historical control data)
- Assessment of human relevance (lack of information on MoA and species differences)

This dissertation seeks solutions to the above challenges with two different foci. On the one hand, the elucidation of species differences between men and laboratory animals and, on the other hand, the analysis of *in silico* models as an alternative to the established *in vivo* studies.

3 Species differences and xenobiotic transporters

3.1 Preamble

The aim of animal testing in a regulatory context is the correct prediction of toxicity, from which the hazard to men can subsequently be estimated. In the case of reproductive toxicology, this is done based on the assumption that the rat and rabbit are good model organisms for all stages of the human reproductive cycle. The great toxicological relevance of species differences first became clear to experts through the thalidomide disaster [31]. Due to the tests carried out at that time, the teratogenic potential of thalidomide was not recognised, which is why the approval requirements and thus also test regulations were subsequently tightened considerably. Thalidomide-induced teratogenicity is species-specific. Rodents are resistant to thalidomide-induced limb malformation, whereas rabbits are sensitive to these effects [95]. To date, the molecular mechanism of thalidomide-induced teratogenicity and thus the cause of species differences is still not completely understood [96].

Species differences in toxicological effects may be due to differences in both mechanism of action and kinetics but have sparsely been studied at these levels. The rabbit in particular has hardly been studied in molecular biology, since unlike the rat it is not usually used as a laboratory animal at universities. The aim of the following publication was to fill some of the knowledge gaps, especially with regard to the rabbit. The focus of the publication was on species differences in the field of toxicokinetics, more specifically on the expression of xenobiotic transporters at the mRNA level. As already explained in section 3.2 ADME, transport across membranes plays a crucial role in the uptake, distribution, and excretion of substances. Xenobiotic transporters are known to transport both endogenous and exogenous substances due to their broad substrate specificity and thus play an important role in the kinetics of xenobiotics. In the study, the expression in liver, kidney and placenta was investigated. Liver and kidney can already play a role in the disposition of xenobiotics in the prenatal phase [12-14], even if their function does not yet correspond to that of the adult stage, and the placenta is known to regulate, among other things, the transport of xenobiotics to the embryo/foetus [11].

In regulatory toxicology, species differences have played a minor role in reprotoxicological studies, which is also due to the fact that very little is known about the background. This applies to the toxic mechanisms but above all to the often-neglected kinetics of the substances. The aim of this paper is therefore to provide a starting point for the systematic investigation of species differences between rats, rabbits, and men in the field of reproductive toxicology, focusing on the expression of xenobiotic transporters that can significantly influence kinetics in the maternal, embryonic and foetal organism.

The following publication was prepared in collaboration with 6 co-authors. All the experimental work, as well as the analysis of the data and the compilation of the first version of the paper, was done by the author of this dissertation. The co-authors were involved in the planning of the experiments, as well as significantly in the review process of the publication.

3.2 Publication 1: Ontogeny of renal, hepatic, and placental expression of ATP-binding cassette and solute carrier transporters in the rat and the rabbit

Full reference:

Weyrich, A., Frericks, M., Eichenlaub, M., Schneider, S., Hofmann, T., Van Cruchten, S., & van Ravenzwaay, B. (2022). Ontogeny of renal, hepatic, and placental expression of ATP-binding cassette and solute carrier transporters in the rat and the rabbit. *Reproductive Toxicology*, 107, 1-9. https://doi.org/10.1016/j.reprotox.2021.10.005

Reproductive Toxicology 107 (2022) 1–9

Contents lists available at ScienceDirect

Reproductive Toxicology

journal homepage: www.elsevier.com/locate/reprotox

Ontogeny of renal, hepatic, and placental expression of ATP-binding cassette and solute carrier transporters in the rat and the rabbit

Anastasia Weyrich [a,*], Markus Frericks [a], Michael Eichenlaub [b], Steffen Schneider [a], Thomas Hofmann [a], Steven Van Cruchten [c], Bennard van Ravenzwaay [a]

[a] Experimental Toxicology and Ecology, BASF SE, Carl-Bosch-Straße 38, 67056, Ludwigshafen, Germany
[b] Bioscience Research, BASF SE, Carl-Bosch-Straße 38, 67056, Ludwigshafen, Germany
[c] Department of Veterinary Sciences, University of Antwerp, Universiteitsplein 1, 2610, Wilrijk, Belgium

ARTICLE INFO

Handling Editor: Dr. Bal-Price Anna

Keywords:
Ontogeny
Rat
Rabbit
ABC transporter
SLC transporter
Placenta
Kidney
Liver

ABSTRACT

Species differences in developmental toxicity can be due to varying expression of xenobiotic transporters. Hence, knowledge on the ontogeny of these transporters, especially in human, rat and rabbit, is pivotal. Two superfamilies of transporters, the ATP-binding cassette (ABC) and the solute carrier (SLC) transporters, are well known for their role in the absorption, distribution and/or elimination of xenobiotics and endogenous substances. The aim of this study was to compare the expression levels of these xenobiotic transporters in liver, kidney and placenta of man, Wistar rat and New Zealand White rabbit during pre- and postnatal development. For this purpose, qPCR experiments were performed for rat and rabbit tissues and the gene expression profiles were compared with literature data from man, rat and rabbit. Data analysis showed large differences in transporter expression in development and between species. These results can be used to better understand developmental toxicity findings in non-clinical species and their relevance for man.

1. Introduction

For the approval of most pesticides and high tonnage chemicals (standard registration of 1000 tonnes or more a year, Annex X of REACH) embryo-foetal developmental toxicity studies in one rodent and one non-rodent species are required by regulatory agencies [1,2]. The preferred laboratory animals used for such studies are rat and rabbit, respectively. As a developmental toxicity finding with a pesticide often hampers marketing of the compound in the EU [3], understanding the mechanism behind the toxicity is critical. When there is proof that the mechanism is specific to the used laboratory animal species, and thus not relevant for man, approval of the compound can still be safeguarded. However, for pesticides, knowledge on the underlying molecular mechanism(s) of developmental toxicity is often lacking [4]. In order to understand the molecular pathways of toxicity, gene and protein data are pivotal. Several data are available for man and rat, but in rabbits there is hardly any information about the functionality of the annotated proteins. This data gap has to be filled in order to determine developmental toxicity mechanisms in this species.

In view of understanding the toxicity of a xenobiotic, kinetics plays a key role. It is already known that the activity of xenobiotic transporters can have a major influence on the disposition and therefore also on the effectiveness of drugs in children and adults [5–9]. A prominent example is the P-gp–dependent toxicity and efficacy profile of opioids in neonates and young infants. At birth the P-gp (MDR1) expression in the brain is limited and increases with postnatal development to reach adult levels at approximately 3–6 months of age. The low expression of P-gp in newborn and young infants could explain their higher sensitivity to opioids compared to adults [7,10]. Several examples of pesticides that interact with xenobiotic transporters are now known as well [11–13]. These transporters can affect pesticide disposition and cause potential toxicity. However, little is currently known about the mechanisms. There are basically three different ways of interaction: (1) pesticides can inhibit transporter activity; (2) pesticides can modulate transporter expression; (3) pesticides can be substrates for transporters, leading to either more or less uptake by cells depending on the role (uptake or efflux) of the transporter [14]. Especially for the third mode of interaction, the ontogeny of transporters can play a role in developmental toxicity and this is therefore the scope of our study. These transporter data can be used in physiology-based pharmacokinetic models to better

* Corresponding author.
E-mail address: anastasia.weyrich@basf.com (A. Weyrich).

https://doi.org/10.1016/j.reprotox.2021.10.005
Received 28 April 2021; Received in revised form 14 September 2021; Accepted 7 October 2021
Available online 30 October 2021

understand and predict potential species differences in exposure (and toxicity) [15–17]. Furthermore, they can help in building AOP frameworks in order to better understand the toxicity mechanism of xenobiotics [18].

This study is intended to better understand potential species-specific differences in developmental toxicity by assessing the basal expression of various xenobiotic transporters in particularly relevant organs for disposition in rats and rabbits during embryo-foetal development. We also included some postnatal stages in order to get a more complete overview on the ontogeny profile, but several other groups have already studied the postnatal expression of transporters in the rat [19–25]. As embryo-foetal development studies are intended for human risk assessment, a comparison was also made with previously published and human data.

The two superfamilies of transporters that are well known for their role in the distribution of xenobiotics are the ATP-binding cassette transporter (ABC) and the solute carrier (SLC) families [9,26–30]. The genes of the ABC and SLC transporter families are homologs and originally arose through gene duplication [31,32]. Most ABC and SLC transporter genes have orthologs in all three regarded species: human, rabbit and rat, but there are exceptions in which genes have none or even several orthologs. This can be seen in Supplementary Table 1, which shows the transporters selected for this study and their characteristics. Furthermore, orthologous genes can differ in their expression and regulation as well as in terms of their substrate affinity (reviewed in [29,33–38]) even if they have a very similar sequence.

In view of embryo-foetal developmental toxicity, the liver and the kidneys of the embryo, the foetus and the dam, as well as the placenta, are important organs for xenobiotic transport. The intestine was not included, as oral absorption is negligible in the embryo-foetal stages. In contrast, the liver and kidney can already play a role in the disposition of xenobiotics during the prenatal phase [39–41], even if their function does not yet correspond to the adult stage. As the placenta is not only crucial for nutrient and oxygen supply to the embryo/foetus and elimination of metabolic waste and carbon dioxide [42], but also regulates the transport of xenobiotics to the embryo/foetus [43], transporter expression was investigated in this organ as well.

Expression data of organic anion transporters, organic anion transporting polypeptides and multidrug resistance proteins during ontogeny in liver and kidney in rats are available [19–25,44–46]. However, most of the rat literature data are restricted to the Sprague-Dawley stock, whereas the Wistar rat is the commonly used stock by pharmaceutical and chemical companies in Europe [47]. The expression of ABC and SLC transporters in man is well characterised in adults (reviewed in [48,49]). Since there is great interest in the ontogeny of transporters in view of paediatrics, there are also several studies and reviews that investigated the transporter expression on mRNA and protein level as well as transporter activity in man at pre- and postnatal time points [6,15,24,50,51]. However, the prenatal time points are limited, as prenatal human material is difficult to obtain. For the placenta most data refer to term placenta and only a few refer to developmental stages, earliest from 16 weeks post conception (wpc) on [17,52–56]. In rats [46,57] and rabbits [58], overall, the data are much more limited. So, the aim of the study was to fill these data gaps and to compare transporter gene expression data between man, rat and rabbit using bioinformatic tools.

2. Materials and methods

2.1. Animals

Female time-mated Wistar HAN rats and female New Zealand White rabbits were obtained from Charles River (Sulzfeld, Germany). Mating of rabbits took place in house. Animals were kept under standard laboratory conditions in fully acclimatized rooms in accordance with the recommendations of the local animal care committee. Temperature range was 22−24 °C in rats and 19−21 °C in rabbits with a relative humidity of 30–70 %. The day/night cycle was 12 h (Light form 06 h–18 h). Air exchange was 15 times per hour. The rats and rabbits were housed individually, and the pups were kept with the dams until sacrifice. Since the animals were not treated and necropsied in deep narcosis, these experiments were not subject of an approval of an ethical committee.

Two rat dams per time point were anaesthetized with isoflurane (CP-Pharma, Burgdorf, Germany) and sacrificed by decapitation at gestation day (GD) 12, 13, 14, 16, 18, 20 and postnatal days (PND) 4 and 21 to collect their offspring. Three rabbit dams per time point were sacrificed with an overdose of Narcoren (Boehringer Ingelheim, Ingelheim, Germany) at GD 12, 19 and 29. At GD16 and GD24 only two dams were used because one was not pregnant and one died, respectively, prior to sacrifice. The sex of the rat and rabbit embryos/foetuses was determined from GD18 and GD24 onwards, respectively. Three male and 3 female animals were used for the following experiments. For the collection of adult tissue, the rat and rabbit dams were used.

2.2. Sample collection

Whole embryo and tissue samples of placenta, liver and kidney (6 embryos/foetuses per time point) used for mRNA extraction were collected, snap frozen, and stored at −80 °C. For a more detailed description of the samples taken, see Table 1.

2.3. mRNA isolation and cDNA synthesis

mRNA was isolated and purified with the Maxwell® RSC simplyRNA Tissue Kit (Promega, Madison, US) according to the manufacturer description. Concentration and purity of mRNA were determined with NanoDrop (Thermofisher, Waltham, US) by measuring absorption at 230, 260 and 280 nm.

cDNA Synthesis was done with GoScript™ Reverse Transcriptase Kit (Promega, Madison, US) according to the manufacturer description. 2500 ng mRNA were deployed per 20 µL reaction and oligo$(dT)_{15}$ primer and random primers were used at the ratio of 3:1. $MgCl_2$ was used at a concentration of 3 mM and 20 U Recombinant RNasin®Ribonuclease Inhibitor was added per reaction. After the reaction was complete, the mixture was diluted 1:10 in nuclease free water for qPCR experiments. cDNA was stored at −20 °C.

2.4. Real-Time qPCR

qPCR was done with GoTaq® qPCR Master Mix (Promega, Madison, US) on LightCycler® 480 Instrument II (Roche Diagnostics, Rotkreuz, Switzerland) according to the manufacturer description. Primers for qPCR were designed so that each one primer span an exon-exon junction and the amplificants were sequenced to check the specificity of the primer pairs. Primer design and alignment were based on Rnor_6.0 (rat) and OryCun2.0 (rabbit) assemblies by NCBI (for primer sequences see Supplemental Table 2 and 3). Additionally, primer efficiency for each primer pair was determined. Forward and reverse primers were used with a final concentration of 500 nM each. The reaction mix also contained 1x GoTaq® qPCR Master Mix and 5 µL cDNA-Sample to a final volume of 20 µL. In every experiment after the standard qPCR reaction protocol (Supplemental Table 4) a melting curve analysis was performed.

2.5. Relative quantification of mRNA

In order to obtain meaningful results from qPCR experiments, the right choice of one or more reference genes is crucial. These genes should have a constant expression level under the experimental conditions. To this end, well-known housekeeping genes were tested for rabbit and rat, and the experimental data were analysed using RefFinder [59]. Glyceraldehyde-3-phosphate dehydrogenase (Gapdh) and peptidylprolyl isomerase A (Ppia) were identified as endogenous reference genes

Table 1
List of organs sampled from rat and rabbit at the different time points. Six biological replicates were taken per time point, except liver in adult rats (n = 4).

Embryonal/Foetal		E	E	E	E	F	F	F			
Rat	GD/PND	GD12	GD13	GD14	GD16	GD18	x	GD20	PND4	PND21	Adult
	Organ	E, P	E, P	E, P	L, K, P	L, K, P	x	L, K, P	L, K	L, K	L, K
Rabbit	GD/PND	x	GD12	x	GD16	GD19	GD24	GD29	x	x	Adult
	Organ	x	E, P	x	L, K, P	L, K, P	L, K, P	L, K, P	x	x	L, K

Abbreviations: GD: gestation day, PND: postnatal day, E: embryo, K: kidney, L: liver, P: placenta.

(ref) for the rat. For rabbit samples phosphoglycerate kinase 1 (PGK1) was used as additional third ref. CP (crossing point) values were determined by LightCycler® 480 Instrument Software (Roche Diagnostics, Rotkreuz, Switzerland) using the Second Derivative Maximum Method [60]. Further analysis were conducted using R [61].

The normalized relative expressions were calculated including the primer efficiency values by means of Pfaffl equation [62]. Thereto the level of mRNA expression in each sample was normalized to the mean value measured for the calibrator sample and was calculated as follows:

$$normrelExp = relExp_{sample} \div relExp_{calibrator}$$
$$= {E_{ref}}^{CP_{sample}} / {E_{target}}^{CP_{sample}} \div {E_{ref}}^{CP_{calibrator}} / {E_{target}}^{CP_{calibrator}}$$

Abbreviations: CP: crossing point; E: Efficiency; ref: reference genes (Gapdh|Ppia, GAPDH|PPIA|PGK1); calibrator: sample with highest relExp at time point adult

In these calculations, sex was not included as a variable because previous studies showed that sex only has an effect on postnatal expression: Hou et al. (2014) and Zhu et al. (2017) examined the expression of Slcos respectively Abccs and Abcg2 in the rat liver during ontogeny and only found gender-specific differences from PND14 onwards [22,25]. Walker et al. (2017) found that there is currently no data showing that the expression of transporters in the human placenta is dependent on the sex of the foetus [56]. Still, in our study we used an equal number of male and female animals at all developmental stages to avoid any sex-bias.

2.6. Statistical analysis of qPCR data

The statistical data analysis for this paper was generated using SAS software, Version 9.4. First log-normal distribution was tested separately for each gene using Shapiro–Wilk test [63] to examine if the assumptions for a Welch *t*-test [64] are at least approximately met. Shapiro-Wilk test was applied to studentised residuals (studentres). The residuals were based on the ANOVA model using the log transformed data. For genes, which are assumed to be not log-normal distributed due to a significant test result (p-values $\leq$ 0.01), outliers (studentres $>$ 3) were eliminated, and log-normal distribution was tested again. For better assessment of the results quantile-quantile plots were created. A pair-wise comparison within the genes between organs and time points was performed via the Welch *t*-test using log-transformed data. The results of Welch *t*-test are shown in the supplemental data (Statistical data)

2.7. Analysis of literature data

In order to have comparative human data and to expand the data of rats and rabbits, additional literature data were evaluated. For this purpose, the very comprehensive RNASeq study from Cardoso-Moreira et al. (2019) [65] was chosen due to its good comparability regarding the time points of sample collection. In this study, expression at many time points of ontogeny in liver and kidney from man (n = 1–4), rabbit (NZW, n = 4) and rat (Holzman Sprague Dawley (SD), n = 4) was analysed. The sex of rats and rabbits was determined at each time point and 2 male and 2 female animals were used. The normalized raw values (RPKM) were used for evaluation, which can be found on webpage http://evodevoapp.kaessmannlab.org. For better comparison one time point per species (human: 7–9 years of age, rabbit: P186-P548, rat: P112) was defined as adult time point and used as normalization factor as in qPCR experiments.

2.8. Plotting data

For interpretation, data were shown in two types of plots. Initially, a heatmap format was used as an overview and for comparisons between genes within species and experiments. Secondly, line plots of each gene were shown to compare expression levels between species. Virtual TPs (Time Points) were used as the x-axis, considering the species-dependent developmental timelines. Corresponding developmental stages can be seen in Supplemental table 5. The figures were produced using the R packages ComplexHeatmap [66] and ggplot2 [67], respectively.

3. Results and discussion

3.1. Transporter expression data in man, rat and rabbit liver, kidney and placenta

Data from our qPCR experiments are shown in Fig. 1. The data from Cardoso-Moreira et al. (2019) are depicted in Supplemental figures 1, 2 and 3. One calibrator sample (marked with "C" in heatmaps) was chosen per gene and experiment and used as normalization factor (see also section 2.5). This nicely illustrates in which organ (liver or kidney) the transporter gene has been expressed mostly in the adult. Some of the selected transporters were particularly expressed in liver or kidney, while others were expressed at similar levels in two or three of the studied organs. Additionally, the relative expression of transporter genes differed distinctly (bar plot with relative expression values of the calibrator sample) from each other and also between species.

3.2. Comparison of transporter expression data in man, rat and rabbit liver and kidney

Line plots with one gene per plot in Figs. 2–4 and Supplemental Figures 4 and 5 clearly show species differences in ontogeny profiles for the xenobiotic transporters. These differences are discussed below.

3.2.1. Differences between the studies and strains

For the analysis of transporter expression in rabbit and rat, our qPCR data were compared with the RNASeq data from Cardoso-Moreira et al. (2019). Although different techniques were used, both experiments used similar time points. For the rabbit, the same strain was used and when comparing the line plots of the two different experiments (Figs. 2–4 and Supplemental Figures 4 and 5), overall the expression curves, such as increase or decrease, and the magnitude of the values, were similar for all genes, albeit differences in curve shapes for some genes. As such, our rabbit data provide a good description of the time course of expression and the magnitude of expression of xenobiotic transporters at mRNA level in this species.

For the rat experiments, two different rat strains were used: Wistar rats for our qPCR experiments and SD rats in the RNASeq experiments by Cardoso-Moreira et al. (2019). These two are the most used strains in research and toxicological studies. The expression curves of both rat

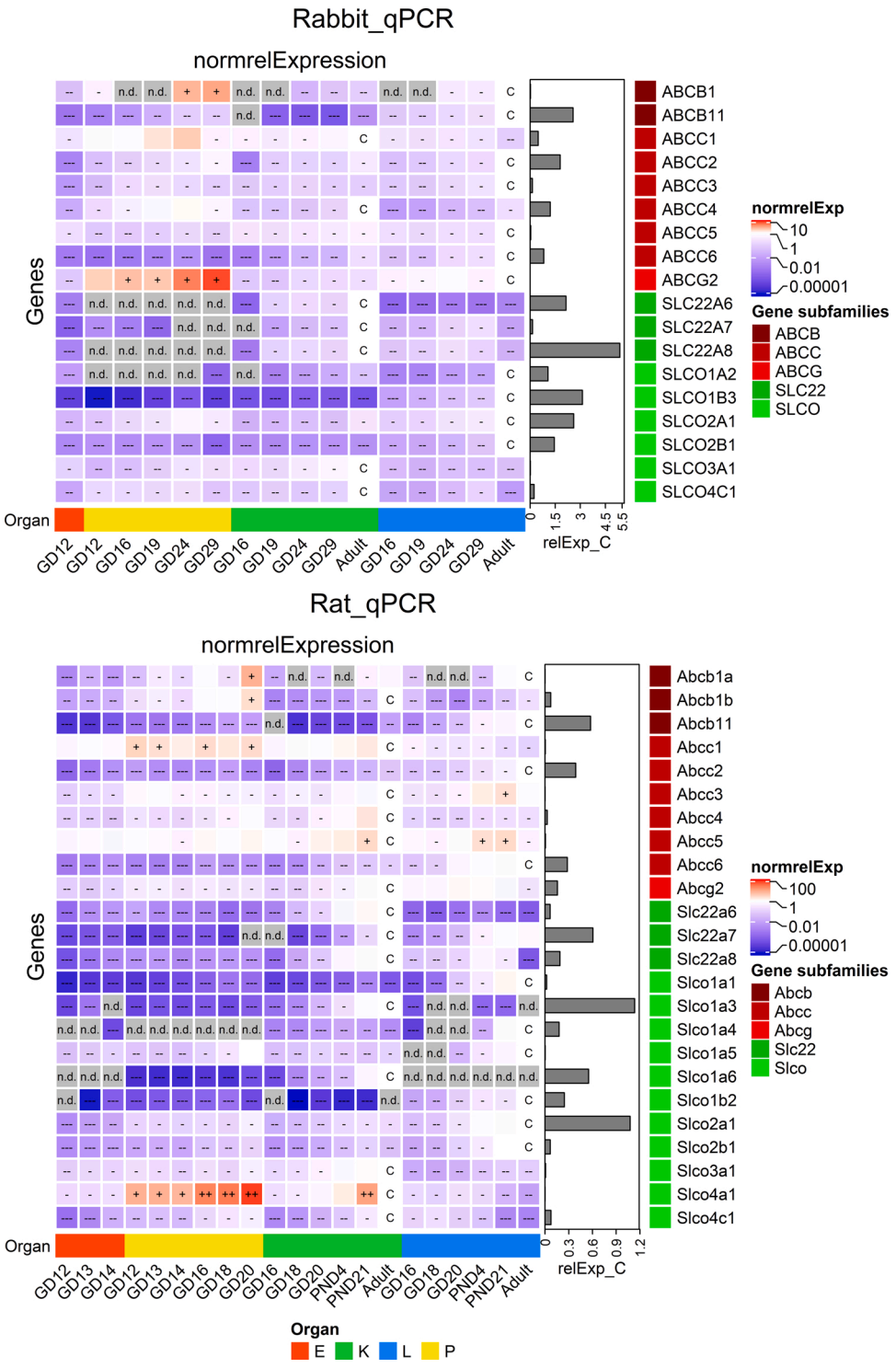


Fig. 1. Heatmaps with normalized relative expression values from qPCR experiments of xenobiotic transporter genes in embryo, placenta, liver and kidney of rabbits and rats at different gestational ages. Measurements were performed in biological replicates of n = 3-6. Normalized relative expression was quantified per gene, using sample (liver or kidney) with highest relative expression at time point adult as normalization factor/calibrator. This sample is marked by "C". The unnormalized values of calibrator samples were shown in bar plot to enable comparison in expression strength between the genes. Meaning of algebraic signs in single cells of the heatmap can be seen in Supplemental table 6. n.d.: not detectable; E: embryo; P: placenta; K: kidney; L: liver, C: calibrator.

strains were very similar which can be seen in the line plots (Figs. 2–4 and Supplemental Figures 4 and 5). This shows that the two strains behave similarly in terms of the expression of the transporters examined, which is important in view of the interpretation of toxicological data in these two different strains.

3.2.2. *Differences in developmental pattern*

Most of the transporters showed an increased expression in the various tissues during ontogeny, although often interrupted by plateau phases during different developmental stages. Some examples were the renal expression of ABCC6 in the rabbit, which did not change significantly between GD24 and GD29 (Supplemental Fig. 4C) and the placental expression of Slco1a5 in the rat, in which the statistically significant increase started only after GD14 (Supplemental figure 5C).

Some transporters also showed higher expression levels during development than in adulthood. For example, the expression levels of human SLCO2A1 and SLCO4A1 in the liver decreased during development. The expression of hepatic rat Slco4c1 raised during the embryonic phase and then significantly decreased with the beginning of foetal phase. Additionally, there are transporters in which expression rose until

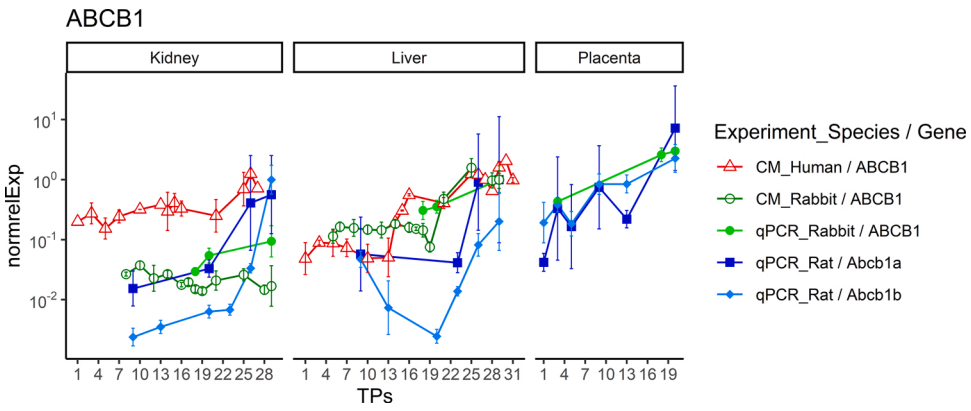


Fig. 2. Normalized relative expression of hABCB1, rabABCB1, ratAbcb1a and ratAbcb1b mRNA in the liver, kidney and placenta of human, rabbit or rat as a function of gestational age. Data was measured by quantitative real-time PCR (qPCR) or by RNASeq experiments by Cardoso-Moreira et al. (2019) (CM). Normalized relative expression was quantified per experiment and gene, using sample (liver or kidney) with highest relative expression at time point adult as normalization factor. qPCR measurements were performed in biological replicates of n = 3-6, while CM measurements worked with n = 4 for rat and rabbit and n = 1-4 for human at each time point. Values are presented as geometric means */ geometric SD.

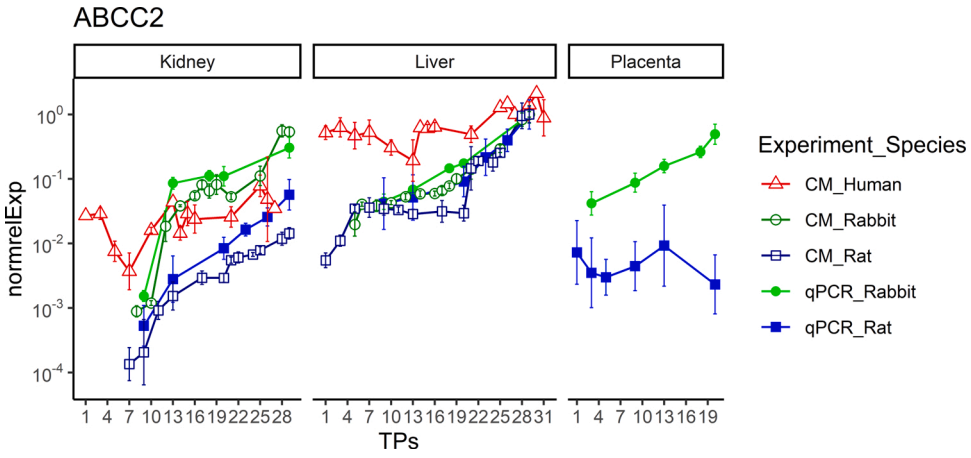


Fig. 3. Normalized relative expression of hABCC2, rabABCC2 and ratAbcc2 mRNA in the liver, kidney and placenta of human, rabbit or rat as a function of gestational age. Data was measured by quantitative real-time PCR (qPCR) or by RNASeq experiments by Cardoso-Moreira et al. (2019) (CM). Normalized relative expression was quantified per experiment and gene, using sample (liver or kidney) with highest relative expression at time point adult as normalization factor. qPCR measurements were performed in biological replicates of n = 3-6, while CM measurements worked with n = 4 for rat and rabbit and n = 1-4 for human at each time point. Values are presented as geometric means */ geometric SD.

postnatal days and then significantly decreased to the adult value, for example hepatic expression of Slc22a8 in the rat (Supplemental figure 5A).

Although the mRNA expression values of transporters do not necessarily relate to their functional activity, the adult value of a functional transporter can provide useful information. For most transporters the adult level of expression is only reached postnatally. However, hepatic human ABCB11 reached already adult level during the foetal phase and renal human ABCC5 showed a constant expression from the earliest developmental stages onwards.

In the following sections, the ontogeny profiles of four transporter genes (ABCB1, ABCC2, SLC22A7 and SLCO2A1) were compared between the species in detail. These transporters were selected because they are characterized by particularly noticeable differences or similarities in their ontogeny profiles between the species (Figs. 2–4).

3.2.3. *ABCB1*

Human ABCB1, also often referred to as p-glycoprotein or multidrug resistance protein 1 (MDR1), is known for its expression in liver and kidney [27]. This is also confirmed by the analysed data (Fig. 2). Expression in liver and kidney was at similar levels in infants compared to adult. The expression during prenatal stages was lower in both organs. Expression levels of hABCB1 in the kidney fluctuated between 0.1 and 0.4 of adult values and increased only after birth. The expression levels in the liver were more than ten times lower than adult levels until time point 13wpc. Afterwards, the expression level increased up to half of the adult level until birth. ABCB1 was expressed significantly differently in rabbit liver and kidney as per the analysis of the qPCR data. In the kidney, expression never reached levels above 0.1 compared to adult liver level. Prenatal expression levels in the rabbit liver were just above 0.1 and increased to the birth level (0.5) only in the second foetal phase. Adult level was reached from time point P14 on. Statistical analysis of qPCR data showed a significant increase from GD24 to adult level. In the rat ABCB1 has two orthologs. Expression differences of Abcb1a in liver and kidney at adult level were insignificantly different. In both organs the adult levels were reached at PND21. Before PND21 expression levels were significantly different and more than ten times lower. Also, Abcb1b was expressed at statistically similar levels in adult liver and kidney, but the ontogeny pattern was different. In the kidney, expression increased significantly after PND4 and was more than 100 times lower at that timepoint than at adult level.

The expression profiles of ABCB1 in the various species differed greatly both in terms of adult expression and developmental pattern. While adult expression of human ABCB1 and rat Abcb1a and Abcb1b was similar in kidney and liver, rabbit ABCB1 showed a significantly lower expression in the kidney than in the liver. Especially in kidney, the human expression levels during development were closer to the adult value than in the rat and rabbit.

3.2.4. *ABCC2*

Human expression of ABCC2, also often referred to as multidrug resistance-associated protein 2 (MRP2), was much higher in liver than in kidney (Fig. 3). In the kidney, expression never reached levels above 0.1 compared to the adult liver. From 4wpc to birth, the hepatic expression of hABCC2 fluctuated around 50 % of the adult level and gradually increased postnatally to adult level. Rabbit ABCC2 was mainly expressed in the liver, but an expression level of twenty percent of the adult liver expression was also achieved in the adult kidneys. Renal expression significantly increased from GD16 to GD19, then expression stayed constant until birth. In the liver, the ABCC2 expression significantly

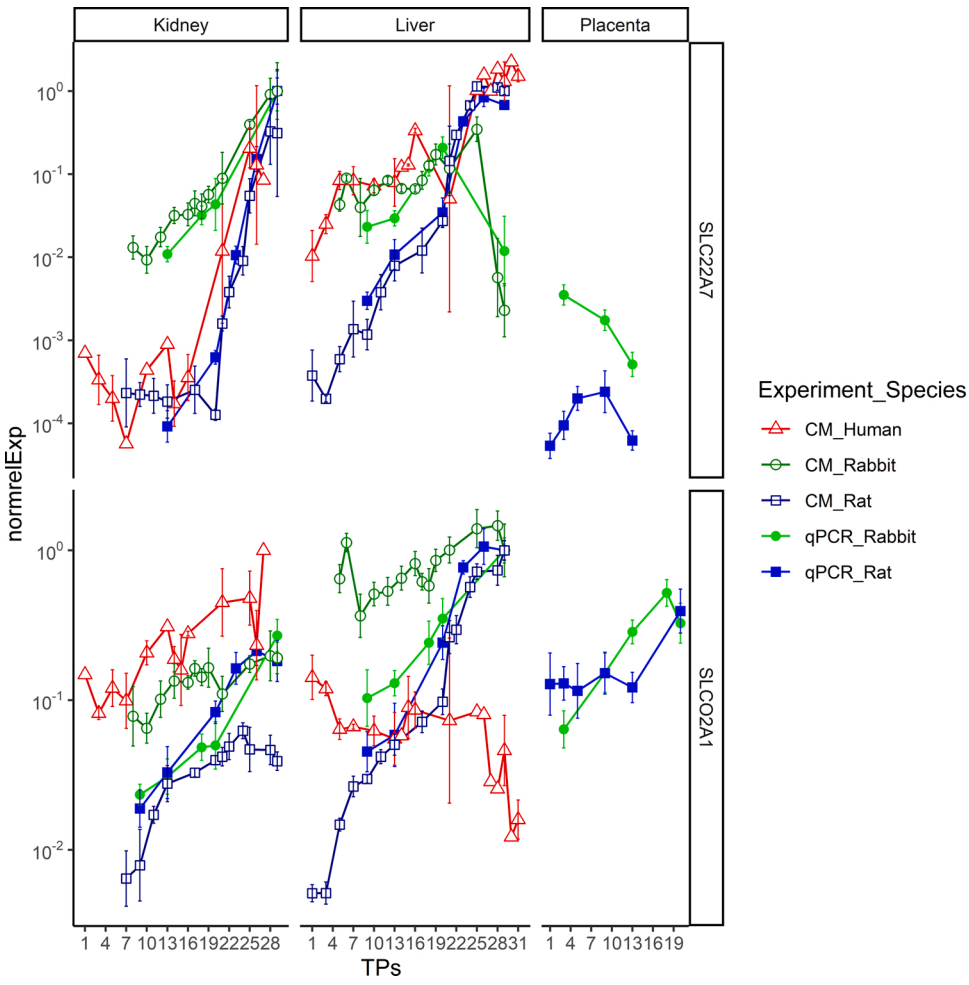


Fig. 4. Normalized relative expression of SLC22A7 and SLCO2A1 mRNA in the liver, kidney and placenta of human, rabbit or rat as a function of gestational age. Data was measured by quantitative real-time PCR (qPCR) or by RNASeq experiments by Cardoso-Moreira et al. (2019) (CM). Normalized relative expression was quantified per experiment and gene, using sample (liver or kidney) with highest relative expression at time point adult as normalization factor. qPCR measurements were performed in biological replicates of n = 3-6, while CM measurements worked with n = 4 for rat and rabbit and n = 1-4 for human at each time point. Values are presented as geometric means */ geometric SD.

increased during development until the adult level was reached at P186, with a constant phase between GD24 and GD29. Rat Abcc2 was significantly more expressed in the liver than the kidney and also expression curve in the liver was very similar to rabbit. Renal expression did not reach expression levels over 0.1 but levels increased significantly during whole ontogeny.

In all species ABCC2 was more expressed in the liver than in the kidney. Interesting to note was that the temporal expression of ABCC2 was very similar in the liver of rabbit and rat and differed to man, in which much higher expression was noted through all stages of development.

3.2.5. *SLC22A7*

Human SLC22A7, also referred to as organic anion transporter 2 (OAT2), was expressed at the highest level in the liver (Fig. 4). Hepatic expression fluctuated during the prenatal phase between 0.01 and 0.3 of adult expression levels, which were reached during infancy. The adult kidney expression was eight times less than the liver and expression increased during the foetal phase. In the rabbit, kidney expression started around 0.01 of the adult level and raised to adult level over the entire developmental period with a statistically constant phase between GD24 and GD29. Liver expression started three times higher than kidney and reached an expression level of 0.35 at time point P14. Then expression decreased significantly, and hepatic adult level was at least 100 times lower than adult kidney expression. Rat Slc22a7 was expressed at significantly different levels in adult kidney and liver. The hepatic adult expression was nearly 70 percent of renal adult expression. Kidney expression was very low (>0.001) during prenatal phase and increased significantly from foetal phase on during whole ontogeny. Liver expression started at similar levels but increased significantly already from the early embryonic time point. Adult liver level had been reached at PND21.

Human SLC22A7 is known for its expression in kidney and liver [35], which was confirmed by the data. This was also true for rat Slc22a7, which expression curve in kidney behaved similar to human. In contrast renal expression of SLC22A7 in the rabbit was much higher through all steps of development. Besides, hepatic expression never reached adult kidney level.

3.2.6. *SLCO2A1*

Expression of SLCO2A1, also known as the prostaglandin transporter, started around 0.1 of adult values in human liver and kidney (Fig. 4). Then expression increased in the kidney up to adult level. In the liver expression decreased 10-fold during ontogeny. In the rabbit, SLCO2A1 was mainly expressed in the liver. Expression in embryonic phase started with values above 0.1 and reached adult level at birth. Expression in adult kidney was three to five times lower than in adult liver and rose one decimal power during development. According to statistical analysis of qPCR data significant increase took place both between GD19 and GD24 and between GD29 and adult time point in kidney and liver. Also rat Slco2a1 was mainly expressed in the liver. Expression significantly increased from GD18 onwards and reached adult level at PND21. Renal expression significantly increased during whole prenatal phase and reached the adult kidney level at PND4.

The expression of SLCO2A1 was totally different between human and the non-clinical species. While rabbit SLCO2A1 and rat Slco2a1 were

mainly expressed in the liver, hSLCO2A1 expression in the liver decreased over time and its main expression was in the kidney.

3.2.7. Differences between species

Using the example of ABCB1, ABCC2, SLC22A7 and SLCO2A1 it was shown that there were large differences in the expression of the transporters between the species and that there were also transporters of which the expression was very similar in two or three of the species examined. In contrast to ABCB1 and ABCC2, which mainly show species differences in their ontogeny pattern, there are large differences in adult expression for SLC22A7 and SLCO2A1. This can lead to differences in the distribution of their respective substrates and thus also have an influence on the potential toxicity of these substrates. However, since the substrate specificity of xenobiotic transporters can partially overlap, other isoforms could compensate for a low transport activity. Furthermore, differences in the translation to protein are possible, which also influences the transporter activity. Therefore, it is critical to assess transporter activity to know whether differences at mRNA expression level affect the distribution of substrates.

Overall, nor the rat nor the rabbit mimics man in terms of the expression of the transporters, as it depends on the transporter of interest. Knowledge on these differences is critical when interpreting toxicity data in view of human safety/risk assessment. In the case of the rat, it does not seem to make any difference with regard to the transporter whether the SD or Wistar strain is used, and the data obtained can be used for both strains.

3.3. Comparison of transporter expression in man, rat and rabbit placenta

In the placenta of rabbit, the expression of ABCB1 (Fig. 2) increased during gestation and was significantly higher than adult level in the kidney from GD24 onwards. Similarly, the expression of the two rat orthologs Abcb1a and Abcb1b in the placenta increased during gestation but significant increase only took place between GD18 and GD20. At GD20 they reached levels above the ones in the adult kidney. These high mRNA levels were also observed by Leazer and Klaassen (2003). In man, the expression of placental ABCB1 has been investigated in numerous studies from gestational week 7 until term [68–70] and expression decreased during gestation.

Expression of rabbit ABCB11 (Supplemental Fig. 4A), also known as the bile salt export pump, significantly increased during gestation until GD24 and reached expression levels of 0.1. Rat Abcb11 mRNA level significantly increased between GD14 and 18 and only reached expression levels of 0.01 similar to adult kidney expression, which is in accordance with the data of Leazer and Klaassen (2003) and St-Pierre et al. (2004). In man, ABCB11 could only be found on mRNA level in placenta samples collected between gestation week 9 and 12 but no expression at term could be measured [71].

In the rabbit placenta, ABCG2 (Supplemental Fig. 4A) (breast cancer resistance protein) mRNA expression increased during gestation from GD19 on and expression was significantly higher than in liver and kidney from GD24 on. Halwachs et al. (2016) showed the functional expression of the ABCG2 efflux transporter in rabbit placenta. The placental expression in the rat was around 0.3 and remained constant during gestation. Expression of ABCG2 in human placenta was shown on protein and mRNA level but the findings about the course of expression during pregnancy are inconsistent [17,53,55].

In the rat, the Abcc transporters (Fig. 3, Supplemental Figs. 4B and 4C), also called multidrug resistance-associated proteins, can be divided into two groups regarding their placental expression. Abcc1, Abcc3, Abcc4 and Abcc5 were expressed in the placenta at similar or even higher levels than in adult liver and kidney, while expression of Abcc2 and Abcc6 was at least 100 times smaller than adult liver expression. Similar results were shown by Leazer and Klaassen (2003) and St-Pierre et al. (2004). The placental expression of ABCC1 in the rabbit was in the same range as adult kidney expression. ABCC2 and ABCC4 showed a significantly increasing expression during gestation up to adult level. Placental expression of ABCC3 and ABCC5 was around 10 times lower and ABCC6 at least 100 times lower than adult expression in liver or kidney. ABCC1−5 are known for their expression in human placenta while expression of ABCC6 has not been demonstrated [17,52,54].

Human SLC22A6−8 are not expressed in the placenta [17]. This is in line with our qPCR data in the rat and rabbit (Supplemental figure 5A). The expression levels were consistently below 0.01 or even below the detection limit. Also, Leazer and Klaassen (2003) measured extremely low expression levels of Slc22a6−8 in the placenta of rats.

SLCO1A2 (Supplemental figures 5B and 5C) is expressed in human placenta [17,56]. In the rabbit only at GD20 a very low expression was measured. The five rodent orthologs were expressed differently in the rat. For Slco1a4 no expression could be measured, whereas expression of Slco1a5 significantly increased from GD14 onwards up to adult level. The mRNA expression of Slco1a1, Slco1a3 and Slco1a6 was very low and maximally reached 0.001. This is in accordance with the results of Leazer and Klaassen (2003) and St-Pierre et al. (2004), who showed that ratSlco1a5 had the highest expression of all genes of the Slco1a subfamily compared to liver and kidney.

Briz et al. (2003) analysed the mRNA expression levels of SLCO1B1 and SLCO1B3 in human placenta. They concluded that SLCO1B3 may play a functional role, whereas expression of SLCO1B1 was very low [72]. In the rabbit only for SLCO1B3 (Supplemental figure 5D) an ortholog was known. Its expression significantly increased during gestation but just reached expression levels around 0.0001. The ortholog in rat is Slco1b2 and its expression in placenta was less than 0.001 during whole gestation.

Rabbit SLCO2A1 and rat Slco2a1 (Fig. 4) were expressed close to adult liver level at GD20. However, expression of rabSLCO2A1 significantly increased during the entire gestation, whereas expression of rat Slco2a1 remained constant until GD18. The expression of SLCO2B1 differed in rat and rabbit. Slco2b1 mRNA levels in the rat discontinuous increased during development and reached levels above 0.1. Expression levels of SLCO2B1 in the rabbit were around 0.005 until GD24 and then significantly decreased. Both SLCO2A1 and SLCO2B1 were expressed in the human placenta whereby expression increased during gestation [53,73].

The expression level of Slco3a1 (Supplemental figure 5E) in the rat placenta fluctuated between 0.1 and 0.4. In the rabbit expression significantly increased from GD16 to GD24 to from 0.03 up to 0.2. SLCO3A1 was also detected in human placenta and expression decreased significantly from first to third trimester [17,71].

In the rabbit only one of the two members of the SLCO4 subfamily is annotated and this is SLCO4C1. For human SLCO4C1 no placenta expression is known but in our qPCR experiments rabbit SLCO4C1 and rat Slco4c1 (Supplemental figure 5 F) both showed expression levels around 0.1 in placenta. In contrast human SLCO4A1 is known for its high placental expression compared to other organs [71,74]. The mRNA expression level of Slco4a1 in rat placenta increased by 10–100 times more than expression in adult kidney. This was in line with the data by Leazer and Klaassen (2003) and St-Pierre et al. (2004).

When comparing the data from rats and rabbits in the placenta, as in the liver and kidney, there were both transporters with a similar and completely different expression profile. A direct comparison with human placenta data is not possible because there are no suitable studies with comparable time points during pregnancy. Above all, our data provide a good starting point for comparing the results from embryo-foetal developmental toxicity studies in rats and rabbits. In addition to the problem of the data situation, structural differences in the placenta between the species also may play an important role in the distribution of xenobiotics [75,76]. These structural differences must be taken into account when evaluating the data from toxicological studies in order to be able to correctly assess the relevance for humans.

4. Conclusion

The aim of this study was to fill data gaps for transporter expression in rat and rabbit and to compare the expression levels of xenobiotic transporters in man, rat and rabbit during development. For this purpose, qPCR experiments for rabbit and rat were performed and additional literature data were analysed. As ontogeny data for xenobiotic transporter expression are rare in the rabbit, this study provides novel data for developmental toxicity studies in this species. Furthermore, this study showed substantial differences in expression of most xenobiotic transporters between the investigated species, which can have an impact on the distribution of xenobiotics and thus on their effects during pre- and postnatal development. These differences with man were found in both rat and rabbit, and as such neither the rat nor the rabbit is a better translational model for human xenobiotic transport in liver, kidney or placenta. The described data should be considered when interpreting developmental toxicity data in rat and rabbit, especially when the affinity of the compound for a specific or several xenobiotic transporter(s) is known.

Conflict of Interest

The authors declare no conflict of interest.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

The authors thank the lab teams of reproductive toxicology and industrial biotechnology for their (technical) assistance and Martina Dammann for her support in statistical data evaluation. This work has been funded by BASF SE.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.reprotox.2021.10.005.

References

[1] EU Law, 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, 2021.

[2] OECD, Test No. 414: prenatal developmental toxicity study. OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, 2018.

[3] EU Law, Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 Concerning the Placing of Plant Protection Products on the Market and Repealing Council Directives 79/117/EEC and 91/414/EEC, 2021.

[4] National Research Council, 4 Mechanisms of Developmental Toxicity, in Scientific Frontiers in Developmental Toxicology and Risk Assessment, The National Academies Press, Washington DC, 2000.

[5] International Transporter Consortium, et al., *Membrane transporters in drug development.* Nature reviews, Drug discovery 9 (3) (2010) 215–236, https://doi.org/10.1038/nrd3028.

[6] K. Brouwer, et al., Human ontogeny of drug transporters: review and recommendations of the pediatric transporter working group, Clin. Pharmacol. Ther. 98 (3) (2015) 266–287, https://doi.org/10.1002/cpt.176.

[7] F. Rodieux, et al., Causes and consequences of variability in drug transporter activity in pediatric drug therapy, J. Clin. Pharmacol. 56 (S7) (2016) S173–S192, https://doi.org/10.1002/jcph.721.

[8] Y. Elmorsi, J. Barber, A. Rostami-Hodjegan, Ontogeny of hepatic drug transporters and relevance to drugs used in pediatrics, Drug Metab. Dispos. 44 (7) (2016) 992–998, https://doi.org/10.1124/dmd.115.067801.

[9] X. Liang, et al., Chapter 10 - Role of transporters in drug disposition and drug-drug interactions, in: S. Ma, S.K. Chowdhury (Eds.), Identification and Quantification of Drugs, Metabolites, Drug Metabolizing Enzymes, and Transporters, second edition, Elsevier, Amsterdam, 2020, pp. 311–337.

[10] J. Lam, et al., The ontogeny of P-glycoprotein in the developing human blood–brain barrier: implication for opioid toxicity in neonates, Pediatr. Res. 78 (4) (2015) 417–421, https://doi.org/10.1038/pr.2015.119.

[11] D.S. Buss, A. Callaghan, Interaction of pesticides with p-glycoprotein and other ABC proteins: a survey of the possible importance to insecticide, herbicide and fungicide resistance, Pestic. Biochem. Physiol. 90 (3) (2008) 141–153, https://doi.org/10.1016/j.pestbp.2007.12.001.

[12] L.A. Clerbaux, et al., Membrane transporter data to support kinetically-informed chemical risk assessment using non-animal methods: scientific and regulatory perspectives, Environ. Int. 126 (2019) 659–671, https://doi.org/10.1016/j.envint.2019.03.003.

[13] O. Fardel, E. Kolasa, M. Le Vee, Environmental chemicals as substrates, inhibitors or inducers of drug transporters: implication for toxicokinetics, toxicity and pharmacokinetics, Expert Opin. Drug Metab. Toxicol. 8 (1) (2012) 29–46, https://doi.org/10.1517/17425255.2012.637918.

[14] N. Guéniche, et al., Implication of human drug transporters to toxicokinetics and toxicity of pesticides, Pest Manag. Sci. 76 (1) (2020) 18–25, https://doi.org/10.1002/ps.5577.

[15] K.W.K. Cheung, et al., Incorporating ontogeny in physiologically based pharmacokinetic modeling to improve pediatric drug development: what we know about developmental changes in membrane transporters, J. Clin. Pharmacol. 59 (Suppl (1)) (2019) S56–s69, https://doi.org/10.1002/jcph.1489.

[16] X. Chu, K. Bleasby, R. Evers, Species differences in drug transporters and implications for translating preclinical findings to humans, Expert Opin. Drug Metab. Toxicol. 9 (3) (2013) 237–252, https://doi.org/10.1517/17425255.2013.741589.

[17] A. Dallmann, et al., Drug transporters expressed in the human placenta and models for studying maternal-fetal drug transfer, J. Clin. Pharmacol. 59 (S1) (2019) S70–S81, https://doi.org/10.1002/jcph.1491.

[18] M. Vinken, et al., Development of an adverse outcome pathway from drug-mediated bile salt export pump inhibition to cholestatic liver injury, Toxicol. Sci. 136 (1) (2013) 97–106, https://doi.org/10.1093/toxsci/kft177.

[19] S.C. Buist, et al., Gender-specific and developmental influences on the expression of rat organic anion transporters, J. Pharmacol. Exp. Ther. 301 (1) (2002) 145–151, https://doi.org/10.1124/jpet.301.1.145.

[20] S. Fattah, P. Augustijns, P. Annaert, Age-dependent activity of the uptake transporters Ntcp and Oatp1b2 in male rat hepatocytes: from birth till adulthood, Drug Metab. Dispos. 43 (1) (2015) 1–8, https://doi.org/10.1124/dmd.114.059212.

[21] B. Gao, et al., Differential expression of bile salt and organic anion transporters in developing rat liver, J. Hepatol. 41 (2) (2004) 201–208, https://doi.org/10.1016/j.jhep.2004.04.029.

[22] W.Y. Hou, et al., Age- and sex-related differences of organic anion-transporting polypeptide gene expression in livers of rats, Toxicol. Appl. Pharmacol. 280 (2) (2014) 370–377, https://doi.org/10.1016/j.taap.2014.08.020.

[23] G. Tomer, et al., Differential developmental regulation of rat liver canalicular membrane transporters Bsep and Mrp2, Pediatr. Res. 53 (2) (2003) 288–294, https://doi.org/10.1203/01.PDR.0000047509.54253.01.

[24] B.D. van Groen, et al., Ontogeny of hepatic transporters and drug-metabolizing enzymes in humans and in nonclinical species, Pharmacol. Rev. 73 (2) (2021) 597–678, https://doi.org/10.1124/pharmrev.120.000071.

[25] Q.N. Zhu, et al., Ontogeny, aging, and gender-related changes in hepatic multidrug resistant protein genes in rats, Life Sci. 170 (2017) 108–114, https://doi.org/10.1016/j.lfs.2016.11.022.

[26] D.G. Waller, A.P. Sampson, 2 - pharmacokinetics, in: D.G. Waller, A.P. Sampson (Eds.), Medical Pharmacology and Therapeutics, fifth edition, Elsevier, 2018, pp. 33–62.

[27] J.M. Scherrmann, 5.04 - the biology and function of transporters, in: J.B. Taylor, D. J. Triggle (Eds.), Comprehensive Medicinal Chemistry II, Elsevier, Oxford, 2007, pp. 51–85.

[28] J.A. Ware, L. Zhang, S.-M. Huang, et al., Chapter 14 - mechanisms and genetics of drug transport, in: A.J. Atkinson (Ed.), Principles of Clinical Pharmacology, third edition, Academic Press, 2012, pp. 217–237.

[29] C.D. Klaassen, L.M. Aleksunes, Xenobiotic, bile acid, and cholesterol transporters: function and regulation, Pharmacol. Rev. 62 (1) (2010) 1–96, https://doi.org/10.1124/pr.109.002014.

[30] J. König, F. Müller, M.F. Fromm, Transporters and drug-Drug interactions: important determinants of drug disposition and effects, Pharmacol. Rev. 65 (3) (2013) 944–966, https://doi.org/10.1124/pr.113.007518.

[31] C.T. Hittinger, S.B. Carroll, Gene duplication and the adaptive evolution of a classic genetic switch, Nature 449 (7163) (2007) 677–681, https://doi.org/10.1038/nature06151.

[32] J. Xiong, et al., Tracing the structural evolution of eukaryotic ATP binding cassette transporter superfamily, Sci. Rep. 5 (2015), https://doi.org/10.1038/srep16724, p. 16724-16724.

[33] É. Bakos, L. Homolya, Portrait of multifaceted transporter, the multidrug resistance-associated protein 1 (MRP1/ABCC1), Pflügers Archiv - European Journal of Physiology 453 (5) (2007) 621–641, https://doi.org/10.1007/s00424-006-0160-8.

[34] B. Hagenbuch, B. Stieger, The SLCO (former SLC21) superfamily of transporters, Mol. Aspects Med. 34 (2-3) (2013) 396–412, https://doi.org/10.1016/j.mam.2012.10.009.

[35] H. Koepsell, The SLC22 family with transporters of organic cations, anions and zwitterions, Mol. Aspects Med. 34 (2-3) (2013) 413–435, https://doi.org/10.1016/j.mam.2012.10.010.

[36] A.T. Nies, D. Keppler, The apical conjugate efflux pump ABCC2 (MRP2), Pflügers Archiv - European Journal of Physiology 453 (5) (2007) 643–659, https://doi.org/10.1007/s00424-006-0109-y.
[37] A.N. Rizwan, G. Burckhardt, Organic anion transporters of the SLC22 family: biopharmaceutical, physiological, and pathological roles, Pharm. Res. 24 (3) (2007) 450–470, https://doi.org/10.1007/s11095-006-9181-4.
[38] M. Roth, A. Obaidat, B. Hagenbuch, OATPs, OATs and OCTs: the organic anion and cation transporters of the SLCO and SLC22A gene superfamilies, Br. J. Pharmacol. 165 (2011) 1260–1287, https://doi.org/10.1111/bph.2011.164.issue-7.
[39] G. Haycock, CHAPTER 2 - NEPHROUROLOGY FROM FETUSHOOD TO ADULTHOOD, in: J.P. Gearhart, R.C. Rink, P.D.E. Mouriquand (Eds.), Pediatric Urology, second edition, W.B. Saunders, Philadelphia, 2010, pp. 11–22.
[40] J.A. Ring, et al., Fetal hepatic drug elimination, Pharmacol. Ther. 84 (3) (1999) 429–445, https://doi.org/10.1016/S0163-7258(99)00046-7.
[41] F.J. Suchy, Functional development of the liver, in: F.J. Suchy, R.J. Sokol, W. F. Balistreri (Eds.), Liver Disease in Children, Cambridge University Press, Cambridge, 2007, pp. 14–27.
[42] L. Donnelly, G. Campling, Functions of the placenta, Anaesth. Intensive Care Med. 15 (3) (2014) 136–139, https://doi.org/10.1016/j.mpaic.2014.01.004.
[43] M. Codaccioni, F. Bois, C. Brochot, Placental transfer of xenobiotics in pregnancy physiologically-based pharmacokinetic models: structure and data, Comput. Toxicol. 12 (2019) 100111, https://doi.org/10.1016/j.comtox.2019.100111.
[44] D.H. Sweet, et al., Organic anion and cation transporter expression and function during embryonic kidney development and in organ culture models, Kidney Int. 69 (5) (2006) 837–845, https://doi.org/10.1038/sj.ki.5000170.
[45] G. Guo, D. Johnson, C.D. Klaassen, Postnatal expression and induction by pregnenolone of the organic anion-transporting polypeptide 2 in rat liver, Drug Metab. Dispos. 30 (2002) 283–288.
[46] M.V. St-Pierre, et al., Temporal expression profiles of organic anion transport proteins in placenta and fetal liver of the rat, Am. J. Physiol. Regul. Integr. Comp. Physiol. 287 (6) (2004) R1505–1516, https://doi.org/10.1152/ajpregu.00279.2003.
[47] J.R. Foster, D. Frost, Chapter 3 - The history of the rat, in: A.W. Suttie (Ed.), Boorman's Pathology of the Rat, second edition, Academic Press, Boston, 2018, pp. 7–12.
[48] L. He, K. Vasiliou, D.W. Nebert, Analysis and update of the human solute carrier (SLC) gene superfamily, Hum. Genomics 3 (2) (2009) 195–206, https://doi.org/10.1186/1479-7364-3-2-195.
[49] V. Vasiliou, K. Vasiliou, D.W. Nebert, Human ATP-binding cassette (ABC) transporter family, Hum. Genomics 3 (3) (2009) 281–290, https://doi.org/10.1186/1479-7364-3-3-281.
[50] K.W.K. Cheung, et al., A comprehensive analysis of ontogeny of renal drug transporters: mRNA analyses, quantitative proteomics, and localization, Clin. Pharmacol. Ther. 106 (5) (2019) 1083–1092, https://doi.org/10.1002/cpt.1516.
[51] B.D. van Groen, et al., Proteomics of human liver membrane transporters: a focus on fetuses and newborn infants, Eur. J. Pharm. Sci. 124 (2018) 217–227, https://doi.org/10.1016/j.ejps.2018.08.042.
[52] J.D. Unadkat, A. Dahlin, S. Vijay, Placental drug transporters, Curr. Drug Metab. 5 (1) (2004) 125–131, https://doi.org/10.2174/1389200043489171.
[53] A.M. Mikheev, et al., Profiling gene expression in human placentae of different gestational ages: an OPRU Network and UW SCOR Study, Reprod. Sci. 15 (9) (2008) 866–877, https://doi.org/10.1177/1933719108322425.
[54] A.A. Joshi, et al., Placental ABC transporters: biological impact and pharmaceutical significance, Pharm. Res. 33 (12) (2016) 2847–2878, https://doi.org/10.1007/s11095-016-2028-8.
[55] K. Vähäkangas, P. Myllynen, Drug transporters in the human blood-placental barrier, Br. J. Pharmacol. 158 (3) (2009) 665–678, https://doi.org/10.1111/j.1476-5381.2009.00336.x.
[56] N. Walker, et al., Placental transporter localization and expression in the Human: the importance of species, sex, and gestational age differencesdagger, Biol. Reprod. 96 (4) (2017) 733–742, https://doi.org/10.1093/biolre/iox012.
[57] T.M. Leazer, C.D. Klaassen, The presence of xenobiotic transporters in rat placenta, Drug Metab. Dispos. 31 (2) (2003) 153–167, https://doi.org/10.1124/dmd.31.2.153.
[58] S. Halwachs, et al., The ABCG2 efflux transporter from rabbit placenta: cloning and functional characterization, Placenta 38 (2016) 8–15, https://doi.org/10.1016/j.placenta.2015.12.005.
[59] F. Xie, et al., miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs, Plant Mol. Biol. 80 (1) (2012) 75–84, https://doi.org/10.1007/s11103-012-9885-2.
[60] Wittwer, C., M. Gutekunst, and S. Lohmann, Method for quantification of an analyte. United States Patent No. US, 1999. 6: p. B1.
[61] R Core Team, R: a Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2019.
[62] M.W. Pfaffl, A new mathematical model for relative quantification in real-time RT-PCR, Nucleic Acids Res. 29 (9) (2001), https://doi.org/10.1093/nar/29.9.e45 p. e45-e45.
[63] S.S. Shapiro, M.B. Wilk, An analysis of variance test for normality (Complete samples), Biometrika 52 (3/4) (1965) 591–611, https://doi.org/10.2307/2333709.
[64] B.L. WELCH, The generalization of 'STUDENT'S' problem when several different population varlances A.R.E. INVOLVED, Biometrika 34 (1-2) (1947) 28–35, https://doi.org/10.1093/biomet/34.1-2.28.
[65] M. Cardoso-Moreira, et al., *Gene expression across mammalian organ development*, Nature 571 (7766) (2019) 505–509, https://doi.org/10.1038/s41586-019-1338-5.
[66] Z. Gu, Complex heatmaps reveal patterns and correlations in multidimensional genomic data, Bioinformatics (2016).
[67] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2016.
[68] S. Gil, et al., P-glycoprotein expression of the human placenta during pregnancy, Placenta 26 (2) (2005) 268–270, https://doi.org/10.1016/j.placenta.2004.05.013.
[69] M. Sun, et al., Expression of the Multidrug Resistance P-Glycoprotein, (ABCB1 glycoprotein) in the Human Placenta Decreases with Advancing Gestation, Placenta 27 (6) (2006) 602–609, https://doi.org/10.1016/j.placenta.2005.05.007.
[70] A.A. Mathias, J. Hitti, J.D. Unadkat, P-glycoprotein and breast cancer resistance protein expression in human placentae of various gestational ages, Am J Physiol Regul Integr Comp Physiol. 289 (4) (2005) R963–R969, https://doi.org/10.1152/ajpregu.00173.2005.
[71] P. Patel, et al., Semi quantitative expression analysis of MDR3, FIC1, BSEP, OATP-A, OATP-C,OATP-D, OATP-E and NTCP gene transcripts in 1st and 3rd trimester human placenta, Placenta 24 (1) (2003) 39–44, https://doi.org/10.1053/plac.2002.0879.
[72] O. Briz, et al., Role of organic anion-transporting polypeptides, OATP-A, OATP-C and OATP-8, in the human placenta-maternal liver tandem excretory pathway for foetal bilirubin, Biochem. J. 371 (Pt 3) (2003) 897–905, https://doi.org/10.1042/bj20030034.
[73] M.E. Kraft, et al., The prostaglandin transporter OATP2A1 is expressed in human ocular tissues and transports the antiglaucoma prostanoid latanoprost, Invest. Ophthalmol. Vis. Sci. 51 (5) (2010) 2504–2511, https://doi.org/10.1167/iovs.09-4290.
[74] M. Nishimura, S. Naito, Tissue-specific mRNA expression profiles of human ATP-binding cassette and solute carrier transporter superfamilies, Drug Metab. Pharmacokinet. 20 (6) (2005) 452–477, https://doi.org/10.2133/dmpk.20.452.
[75] A.M. Carter, A.C. Enders, Placentation in mammals: definitive placenta, yolk sac, and paraplacenta, Theriogenology 86 (1) (2016) 278–287, https://doi.org/10.1016/j.theriogenology.2016.04.041.
[76] A.M. Carter, A.M. Mess, Mammalian placentation: implications for animal models, in: L.M. McManus, R.N. Mitchell (Eds.), Pathobiology of Human Disease, Academic Press, San Diego, 2014, pp. 2423–2442.

3.3 Outlook

Species differences of xenobiotic transporters can be considered on several levels: DNA, mRNA, protein, and functionality/activity. As shown in colour in figure 4, the state of knowledge differs at the individual levels: While species differences have already been systematically researched at the DNA and mRNA level (green), there is still a need to catch up at the protein level and in functionality/activity (red). This is explained in detail in the following sections.

At the DNA level, the transporters are annotated in all three species (GRCh38.p14 (men), Rnor_6.0 (rat) and OryCun2.0 (rabbit)). Here, the first species differences already become visible: when comparing the genes of the xenobiotic transporters of men with those of rats and rabbits, orthologues are annotated for most genes in all three species. However, there are also genes that have no orthologues in any of the species (SLCO1B1 is missing in rabbits) and genes that have several orthologues in one of the species (5 rodent orthologues of SLCO1A2).

The species differences at the mRNA level were systematically investigated in the presented paper, thus closing this data gap. This showed that there are large differences in the expression of the transporters during ontogenesis between the species.

The next level is the investigation of expression at the protein level. For some transporters, data are already available in men and rats using different methods such as western blotting, immunofluorescence microscopy and targeted proteomics [97-100]. Subcellular trafficking plays a major role in the post-translational regulation of transport proteins, as a transporter can only be functional if it is located in the membrane [101-103]. A great advantage of immunofluorescence microscopy is therefore that the localisation of the transport proteins can be assessed: here it is of relevance on one hand whether the transporter is present in the cytoplasm or the membrane and on the other hand whether it is present at the apical or basolateral membrane. By isolating subfractions of the plasma membrane (apical and basolateral domains), this distinction is possible with western blotting [104, 105].

Determining the functionality and activity of xenobiotic transporters is generally difficult due to the low substrate specificity, as many substances can be transported by several transporters and most transporters have very many substrates [102, 106]. There are several *in vitro*, *ex vivo* and *in vivo* methods to study the function of xenobiotic transporters [107]. Among the *in vitro* methods, cell-based assays based on primary cell cultures, sandwich cultures of primary hepatocytes; immortalised cell lines or cell lines transfected with transporters play a major role [99, 108-112]. *Ex vivo* models include isolated and perfused organs or tissues such as the human placental perfusion method [113]. *In vivo*, the properties of xenobiotic transporters can be studied by transporter gene knockout models in mice [114].

For the correct and complete assessment of species differences during ontogeny, knowledge about the expression, regulation, functionality, and activity of xenobiotic transporters is

essential. This is particularly important for the rabbit, for which it is hardly known whether the annotated genes lead to functional transporters at all. The investigation of expression at the protein level would be technically possible in rats and rabbits analogous to the presented paper with the help of Western blotting, immunofluorescence microscopy and targeted proteomics, but would still be associated with considerably higher time and financial expenditure. In men, there is always the problem of obtaining samples for ethical reasons, especially at the prenatal time points, and the samples are also subject to great variability. As described in the previous section, it is possible to investigate the activity of transporters, with current research focusing on *in vitro* and *ex vivo* methods. The extent to which the data obtained from these methods correspond to the *in vivo* situation depends on the model, since protein expression within the cells changes, sometimes massively, after collection [25]. Especially for prenatal developmental stages, however, there is no activity data of transporters so far and thus a great need for research.

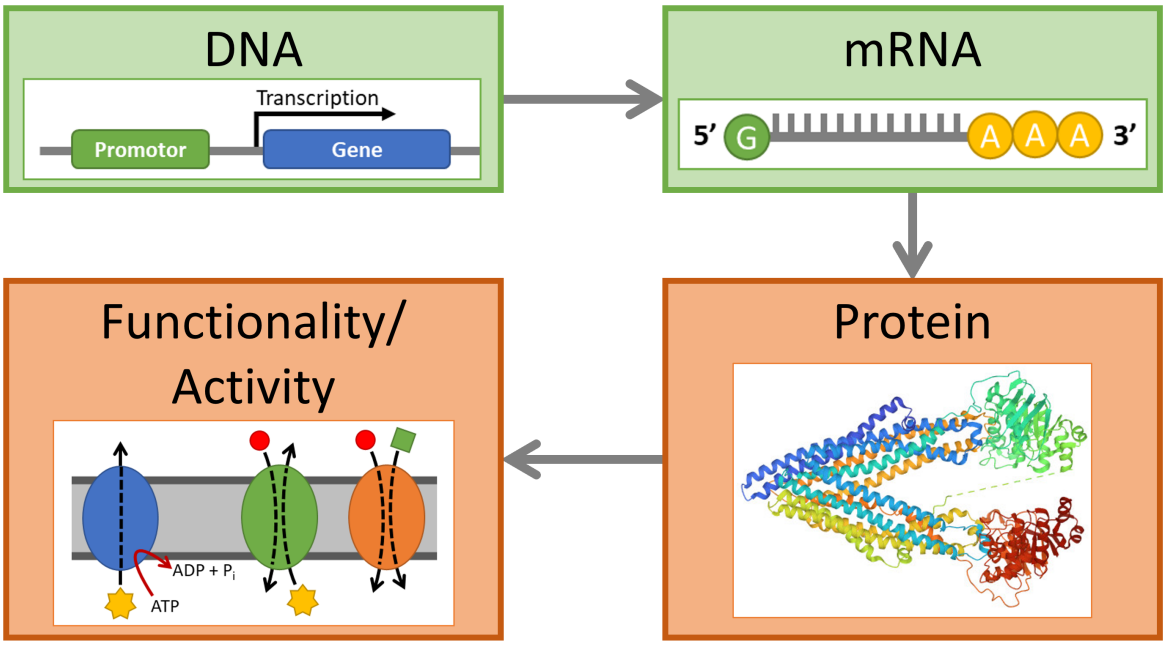


Figure 4: State of knowledge for species differences of xenobiotic transporters on different levels: While species differences have already been systematically researched at the DNA and mRNA level (green), there is still a need learn more at the protein level and in functionality/activity (red). The structure shown in the protein box is that of the P-glycoprotein published by Aller et al. in 2009 and available via RCSB PDB - 3G5U [115].

4 *In silico* prediction models for reprotoxicity

4.1 Preamble

In 2010, the EU adopted Directive 2010/63/EU on the protection of animals used for scientific purposes. This defines the clear goal of completely replacing studies with live animals for scientific and educational purposes as soon as this is scientifically possible [116]. This goes hand in hand with increased support for the development of alternative approaches. It is clear from this guideline that the objective has evolved from the 3Rs Principle to the complete replacement of animal testing.

From the perspective of reproductive toxicology, the search for alternatives to animal testing makes sense not only for ethical reasons, but also for other reasons. Reprotox studies are time and cost intensive, allow only a low throughput, and require a highly skilled laboratory team and study director. In addition to *in vitro* methods, *in silico* methods for predicting reprotoxicity are therefore becoming increasingly important. They are mainly used in screening but are now already required by the authorities for the toxicological assessment of the potential reprotoxicity of pesticide metabolites and contaminants.

The following publication evaluated available predictive models for reproductive toxicology using a pesticide database. The strengths and weaknesses of the models were analysed and suggestions for improving the models were developed on this basis. As the successful development and use of *in silico* prediction models depends on the collaboration of *in silico* experts, regulatory toxicologists, and reproductive toxicologists, all three groups of experts were addressed within the publication. The aim was to raise awareness of the specific issues in reproductive toxicology, to draw attention to the challenges in evaluating models and to highlight data requirements for a successful model.

The following publication was produced in collaboration with 4 co-authors. The statistical evaluation of the models was done by the author of this dissertation, graphically processed and the relevant passages written in cooperation with Madeleine Joel. The introduction as well as the review process was a co-production of all authors.

4.2 Publication 2: Review of the state of science and evaluation of currently available *in silico* prediction models for reproductive and developmental toxicity – a case study on pesticides

Full reference:

Weyrich, A., Joel, M., Lewin, G., Hofmann, T., & Frericks, M. (2022). Review of the state of science and evaluation of currently available *in silico* prediction models for reproductive and developmental toxicity: A case study on pesticides. *Birth Defects Research*, 1– 31. https://doi.org/https://doi.org/10.1002/bdr2.2062

Received: 11 March 2022 | Revised: 10 May 2022 | Accepted: 28 May 2022

DOI: 10.1002/bdr2.2062

RESEARCH ARTICLE

Birth Defects Research | Society for Birth Defects Research & Prevention | WILEY

Review of the state of science and evaluation of currently available *in silico* prediction models for reproductive and developmental toxicity: A case study on pesticides

Anastasia Weyrich[1] | **Madeleine Joel**[2] | **Geertje Lewin**[2] | **Thomas Hofmann**[1] | **Markus Frericks**[3]

[1]Experimental Toxicology and Ecology, BASF SE, Ludwigshafen, Germany

[2]Preclinical Science – Föll, Mecklenburg & Partner GmbH, Münster, Germany

[3]Agricultural Solutions – Toxicology CP, BASF SE, Limburgerhof, Germany

Correspondence
Anastasia Weyrich, Experimental Toxicology and Ecology, BASF SE, Carl-Bosch-Straße 38, 67056, Ludwigshafen, Germany.
Email: anastasia.weyrich@basf.com

Abstract

Background: *In silico* methods for toxicity prediction have increased significantly in recent years due to the 3Rs principle. This also applies to predicting reproductive toxicology, which is one of the most critical factors in pesticide approval. The widely used quantitative structure–activity relationship (QSAR) models use experimental toxicity data to create a model that relates experimentally observed toxicity to molecular structures to predict toxicity. Aim of the study was to evaluate the available prediction models for developmental and reproductive toxicity regarding their strengths and weaknesses in a pesticide database.

Methods: The reproductive toxicity of 315 pesticides, which have a GHS classification by ECHA, was compared with the prediction of different *in silico* models: VEGA, OECD (Q)SAR Toolbox, Leadscope Model Applier, and CASE Ultra by MultiCASE.

Results: In all models, a large proportion (up to 77%) of all pesticides were outside the chemical space of the model. Analysis of the prediction of remaining pesticides revealed a balanced accuracy of the models between 0.48 and 0.66.

Conclusion: Overall, predictions were only meaningful in rare cases and therefore always require evaluation by an expert. The critical factors were the underlying data and determination of molecular similarity, which offer great potential for improvement.

KEYWORDS

in silico predictions, *in silico* protocols, pesticide, QSAR, reproductive toxicology

1 | INTRODUCTION

Reproductive toxicity (Reprotoxicity) is one of the most critical factors in pesticide approval. Due to the 3R principle, the approval authorities are demanding more and more *in silico* evaluations for assessing reprotoxicity. Several models are available using generalized positive or negative calls not evaluating the particular endpoint or

study design. This paper aims at discussing the difficulty and relevant parameters in designing adequate *in silico* models for developmental and reproductive toxicology. To illustrate the difficulties, available models have been tested using a database of 310 pesticides, which are data rich and where testing follows the OECD testing guidelines.

1.1 | Complexity of reproductive toxicology

Reproductive toxicology (Reprotox) reflects the entire circle from formation and maturation of gametes through mating and conception, the embryonic and fetal development, postnatal adaptations, up to sexual maturation of the offspring. Due to the multitude of processes, pathophysiological disturbances may be observed as functional (e.g., altered estrous cyclicity, impaired reflex ontogeny) or structural (e.g., malformations, delayed bone ossification) anomalies or as behavioral alterations (e.g., missing mating drive, altered maternal behavior). Assessing all these factors in one study takes a very long time (approximately two years in rodents) and anomalies sometimes can hardly be appointed to a single interference. Therefore, the entire reproductive cycle is often broken down to several sections, each being tested separately. In this way, particular aspects can be assessed in more detail. However, the enormous animal consumption, time, and cost remain, and it is a great incentive for the development of alternative *in vitro* and *in silico* methods.

In the following section, the *in vivo* studies are presented based on their OECD guidelines, which must be carried out for the registration of a pesticide in the EU. Since these studies are a potential data basis for *in silico* models, knowledge of the assessed endpoints and their classification in the overall toxicological context is of great importance.

1.1.1 | Current OECD Guidelines for assessment of reprotoxicity

In the following section, the major study types used for the generation of reproductive and developmental toxicity data for pesticides and chemicals are outlined. The usual species used in these studies are rats as rodents and rabbits as non-rodents. Additional study types like OECD 422, developmental neurotoxicity studies, or pharma study types are not covered but are equally important and contribute to the available database.

OECD 414: Prenatal Developmental Toxicity Study in one rodent and one non-rodent (OECD, 2018a)

Young mature nulliparous rats are used. Animals in the estrous phase are mated overnight with a male. Successful mating is detected by sperms in the vaginal lavage and defines gestation day 0. Estrous phase in rabbits can be detected by reddening of the vulva if provoked by estrogen injection. Rabbits are mated with a male of proven fertility. Mating is confirmed by the presence of spermatozoa in the vaginal lavage. Alternatively, artificial insemination after hormone treatment can be performed in rabbits. Ovulation occurs approximately 10 hr after mating or estrogen injection.

Animals are then allocated to the different treatment groups (one control group, three treatment groups). Usually, each group consists of 22–24 animals to generate 20 litters per group. Treatment begins at implantation (Days 5–6 in rats and Day 6 in rabbits) and continues until the day before scheduled sacrifice. On Day 20/21 (rats) and 28/29 (rabbits), dependent on strain and/or laboratory, the animals are delivered by cesarean section. Cesarean section is done since otherwise malformed born pups would be lost by cannibalism.

During treatment, the behavior of the animals is carefully observed. Body weight and food consumption are recorded at regular intervals. At cesarean section, the uterus is opened. The numbers of corpora lutea, implantations, early and late resorptions as well as live and dead fetuses are determined. Individual fetal body weights are recorded. All fetuses are examined for external abnormalities. In rats, one half of the fetuses is examined for visceral alterations. The other fetuses are eviscerated, skinned, and evaluated for skeletal alterations, which is usually done by staining with Alcian Blue (cartilage) and Alizarin Red (bones). In rabbits, all fetuses are examined for visceral examination and then eviscerated, stained with Alcian Blue and Alizarin Red and examined for skeletal examination. External, visceral, and skeletal findings are usually classified as malformations (a permanent structural change that is likely to adversely affect the survival or health of the species under investigation) and variations (a change that occurs within the normal population under investigation and is unlikely to adversely affect survival or health).

There are two different study types available for assessing reprotoxicity, which are explained in the next section:

OECD 416: Two-Generation Reproduction Toxicity Study in rodents (usually rats) (OECD, 2001)

The objective of this study is the determination of potential effects on maturation of gametes, mating,

fertilization, pre-implantation stages, implantation. Further potential adverse effects encompass estrous cycle, transport of the fertilized egg, pregnancy, birth, lactation, and growth of the offspring across two generations. In males, effects on libido and epididymal sperm maturation are possible, which cannot be detected otherwise.

Groups of 25 male and 25 virgin female rats, 5–9 weeks of age, are used in this study and allocated randomly to the treatment groups (one control group, three treatment groups). The animals are treated for 70 consecutive days (56 days in case of mice) prior to mating until sacrifice. This time covers a whole spermatogenic cycle including sperm maturation in the epididymis. After evidence of mating, that is, presence of spermatozoa in vaginal smears in the morning, the females are separated from the assigned male and allowed to deliver their F1 offspring. Standardization of offspring at postnatal day (PND) 4 is optional. After 3 weeks of lactation, the F1 animals are separated from their mothers, which are then euthanized. The uterus is opened, and the number of implantation scars is counted. Dosing of F1 animals is then continued for at least 10 weeks before they are mated. After evidence of mating, the females are separated and allowed to deliver their F2 offspring. Standardization of offspring at PND 4 is also optional here. After 3 weeks of lactation, the F2 animals and the maternal animals are euthanized. Males are euthanized when the mating outcome is sufficient. Reproductive organs are weighed and examined histopathologically.

Examined parameters consist of body weight, food consumption, estrous cycle determination, litter parameters, anogenital distance, developmental landmarks of F1 offspring (e.g., preputial separation, vaginal opening), and spermatological examinations. For this purpose, sperm samples are taken from the cauda epididymis and assessed for sperm concentration and sperm motility (motionless, locally motile and progressively motile). Alternatively, computer-assisted sperm analysis (CASA) can be used. Additionally, testicular spermatid head count is determined after homogenization of the testis. Sperm morphology is evaluated by assessment of abnormal head, mid-piece, and tail.

Recently, EFSA required the assessment of nipple retention in male pups around PND 14.

OECD 443: Extended One-Generation Reproductive Toxicity Study in rodents (usually rats) (OECD, 2018b)

This study design originally has been discussed as a replacement for the Two-Generation Reproduction Toxicity Study design, as it requires considerably less animals. The study design is similar to the Two-Generation Reproduction Toxicity Study, but ideally covers only the F1 generation. Optionally developmental neurotoxicity cohorts and a developmental immunotoxicity can be added in the F1 generation.

Groups of 25 male and 25 virgin female rats are used in this study and allocated randomly to the different treatment groups. Pre-mating treatment is at least 2 weeks in males and females. In practice, sometimes a 10-week pre-mating treatment is required by authorities. The age of the animals depends on the pre-mating period (10 weeks treatment: 5–6 weeks old; 2 weeks treatment: 11–12 weeks old). After evidence of mating, the females are separated from the assigned male and allowed to deliver their F1 offspring. Anogenital distance in both sexes and nipple retention are assessed in males. Standardization of offspring at PND 4 is optional. After total 10 weeks of treatment, hematological and clinical chemistry examinations, urinalysis, assessment of organ weights, and histological examination of numerous organs are carried out in the parental animals. Spermatological examinations as described for the OECD 416 study are performed. At weaning, offspring are assigned to the following cohorts:

F1-1A (Reprotoxicity): These animals (20 females and 20 males) are dosed daily from PND 22 and euthanized at the age of 13 weeks and examined in the same way as the parental generation.

F1-1B (Reprotoxicity): These animals (20 females and 20 males) are dosed daily from PND 22 and euthanized at the age of 14 weeks. Reproductive organs and a limited number of other organs are weighed and preserved for possible histopathological examination. If there is evidence of a change of reproductive parameters in the F1A cohort, which warrants further data, these animals are used for breeding and generation of a F2 generation, which is raised and examined like the F1 offspring. It should be noted (although not mentioned in the guideline) that in this case 20 F2 litters should be produced. Therefore, it may be prudent to increase the size of the F2 generation to 25 males and 25 females and consequently increase also the number of parental animals.

F1-2A (optional) Developmental Neurotoxicity: These animals (10 males and 10 females; one male or 1 female out of 20 litters) are subjected to detailed neurological examinations (functional observation battery, motor activity). They are euthanized in Weeks 11 and 12 by perfusion fixation. Central and parts of the peripheral nervous system are preserved, fixed, and embedded in paraplat or plastic (epoxy resin) and histologically examined.

F1-2B (optional) Developmental neurotoxicity: These animals (10 males and 10 females; one male or 1 female out of 20 litters) are euthanized on PND 22, undergo perfusion fixation and are used for assessment of brain weight

and histological examination of brain and brain-associated structures.

F1-3 (optional) Developmental immunotoxicity: These animals (10 males and 10 females; one male or 1 female out of 20 litters) are used at PND 56 ± 3 in a T-cell-dependent antibody response assay (TDAR), for example, the primary IgM antibody response to a T-cell-dependent antigen, such as Sheep Red Blood Cells or Keyhole Limpet Hemocyanin. Additional pups may be required from the control group to act as positive control animals in TDAR. The response is evaluated by counting specific plaque-forming cells in the spleen or by determining the titer of SRBC- or KLH-specific IgM antibody in the serum by ELISA, at the peak of the response.

Alternative study designs

Many alternative *in vivo* non-mammalian and *in vitro* approaches to contribute to the 3Rs concept (Russell & Burch, 1959) exist, but none is accepted by regulatory agencies as alternative test system for registration of pesticides. The major drawback of these alternatives is that the interaction with the maternal compartment is missing.

However, since these methods are of great interest in current research, examples are mentioned for the sake of completeness:

- the Zebra fish embryotoxicity test (ZET) (Selderslaghs, Van Rompay, De Coen, & Witters, 2009)
- the frog embryo teratogenesis assay Xenopus (FETAX) (Bantle, Fort, & James, 1989)
- the whole embryo culture test (WEC) (Piersma et al., 2004)
- the embryonic stem cell test (EST) (Seiler & Spielmann, 2011)

1.1.2 | Differences between study guidelines

One aspect that is often not considered in the comparison of toxicity studies is the change in the underlying experimental guidelines. In case of reprotoxicity, this can have tremendous impact. For example, the original versions of the OECD 414: Developmental Toxicity Testing guideline required dosing only during embryogenesis and organogenesis. In the rat, this is between gestation days (GD) 6–15; in the rabbit 6–19. In the more recent guideline, this was adapted to also cover the later intrauterine maturation leading to treatment between GD 6–20 in the rat and 6–29 in the rabbit. In both guideline versions, the animals were euthanized and delivered by cesarian section to achieve a standardized read out. By use of the older study design developmental delays, for example, ossification effects might have recovered by the last day of pregnancy. This "recovery" period is not present in the newer test design. Furthermore, the day of cesarean section varies between GD 20 and 21 in rats and GD 29 and 30 in rabbits among laboratories and between animal strains. Especially in rats, this difference has significant impact on the ossification status and fetal weight. Since dosing is based on the dam's weight and fetal weight becomes a significant part of this in the last stage of gestation, the high dose tolerated by the dams is expected to be lower in the new study design in many cases.

Another change in the guideline with strong impact is that in older studies, often only bone was stained. In more recent experiments, a co-staining for cartilage is often applied, which allows a much better, standardized analysis of ossification effects.

The impact of guideline changes is even more prominent in OECD 416 (2-generation study). A significant array of additional parameters has been added. Many of these are related to sexual maturation and endocrine disruption, such as anogenital distance, nipple retention, vaginal opening, and preputial separation. More recent changes include measurement of thyroid hormones. Therefore, results for these endpoints are not available for historic studies performed according to the old OECD protocols. In the meantime, depending on the regulatory framework, the OECD 416 is often replaced by the OECD 443. The assessed reprotoxic endpoints in both studies are similar, but in the OECD 443 the F2 generation is often avoided unless triggers are calling for generation of the F2. Additionally, the OECD 443 can contain cohorts for the assessment of developmental neurotoxicity and immunotoxicity.

Additional critical parameters are dose setting, which in historic times often used large spacings, for example, 100, 500, and 5,000 ppm. If the top dose showed excessive toxicity, and the low dose displayed no effects, the extend of effects on reproductive performance or sexual maturation cannot be clearly defined. For NOAEL setting this is not a problem, but for hazard and risk assessment, and also for QSAR, the widely spaced dose setting can mask effects at lower toxicity levels.

In addition to the OECD guidelines just presented, which are used for the classification of pesticides and chemicals in the EU, there are further guidelines for assessing reprotoxicity:

- US EPA OPPTS for the risk assessment of chemicals and pesticides
- ICH-Guidelines for risk assessment for drug authorization, used in the EU (EMA), US (FDA), and Japan (MHLW)

These guidelines display great similarity in their general structure, but there are also some differences about the exposure period and the points considered.

1.1.3 | Causes and mechanisms of reprotoxicity

ADME during gestation and lactation

Critical points in reprotox that have not been adequately explored are ADME and metabolism. The fetuses and pups have different exposure conditions (De Schaepdrijver, Annaert, & Chen, 2019). In utero, exposure to the fetus is largely mediated by maternal supply via the placenta (Tetro, Moushaev, Rubinchik-Stern, & Eyal, 2018). The exposure will therefore be restricted to bioavailable active ingredients and their bioavailable metabolites. Usually, metabolism information is only available for non-pregnant animals. Due to the physiological changes in pregnancy, ADME parameters between pregnant and non-pregnant animals can be significantly different, which can lead to an unknown pattern of exposure (Avram, 2020; Tasnif, Morado, & Hebert, 2016). In order to bundle knowledge about the metabolism of pesticides, EFSA initiated the creation of MetaPath with the EU transparency regulation, which can be used in the future, among other things, for the development of PBPK models.

Placental transfer can be a limiting factor for distribution, since the placenta is designed to form a protective barrier protecting the fetus from xenobiotic compounds. A number of models for placental transport have been proposed and can potentially contribute to an assessment of the fetal exposure situation.

The exposure of the offspring is initially via the meconium, maternal skin contacts and if the compound is fat soluble via milk. Only after pups start ingesting food, approximately around Days 10–14, does dietary exposure become a dominant factor. Currently, no systematic database for milk transfer is available across pharmaceuticals, pesticides, and chemicals. Therefore, log*p* values are a logical way of approximation. Here again, the metabolism of parent and data on tissue distribution into fat should be taken into account, which has not been systematically collected. Having a respective database would be a valuable addition into the toolset of PBPK models to evaluate.

An additional important parameter is the difference in the expression and activity of phase I—III enzymes. Fetal and pup metabolism and excretion is often limited while immature. For example, most transporters only reach maximal expression at around PND 21 in both liver and kidney. While significant data on the rat are available for a number of phase I—III enzymes, the database for humans and rabbits, as the second relevant species for teratogenicity testing, is limited (De Schaepdrijver et al., 2019).

All in all, too little is known about the exact ADME of pesticides during gestation or lactation. Whether the pesticides cross the placental barrier are metabolized by the fetus, exposure takes place via the milk or how ADME works in the pup are questions that cannot be answered for most pesticides. To make matters worse, the embryo/fetus/pup changes over the entire period under consideration, which is why it can be assumed that this also applies to ADME.

Importance of maternal toxicity and species differences

Based on the complexity of reprotoxicity, a vast interplay with related areas such as pathology, endocrinology, and general toxicology is necessary. To make the matter even more complex, an interrelation between generation effects can be seen, such as impaired maternal nutritional status leading to lower numbers of follicles maturing, lower reproductive success, and subsequent lower numbers of live born pups (Khera, 1987; Nitzsche, 2017; Theunissen et al., 2016). Or maternal toxicity can lead to a less than optimal uterine environment, lower nutritional supply to the fetus, possibly resulting in lower fetal weight and delayed skeletal ossification. The inherent role of maternal toxicity has gained increasing attention in the last decade as it assists data interpretation.

Furthermore, different animal models respond differently to exogenous stress factors. While rabbits for example often react with abortions, rats tend to maintain their pregnancies but may display higher numbers of resorptions, lower fetal weight, and developmental delay in their offspring. An additional factor, which is often overlooked, is the documentation of negative results, parameter that was assessed but is not affected. In the future, not only the documentation but also the publication of those can help to shed light on affected pathways.

Therefore, for each reprotoxicity assessment, the right time frame and route of exposure, the most appropriate animal model and a well-suited laboratory with sufficient experience have to be carefully selected.

Adverse outcome pathways

While the conservative study approach can connect between an exposure at a certain timepoint and an outcome, it usually gives no clear information on the mode of action of adverse outcomes. For this, a different approach was developed.

In order to sort parameters, connect cause and consequences and subsequently organize scientific knowledge, the conceptual framework of Adverse Outcome Pathways (AOPs) was initiated. They are intended to aggregate

knowledge currently dispersed in various sources from case studies, journal articles to databases into a systematic and accessible format that facilitates use of that knowledge.

AOPs are based of several principles:

- Linking a molecular initiating event (MIE) via several key events to an adverse outcome.
- Modular AOPs can assemble into AOP networks.
- And AOPs are living documents, reflecting the current state of science and open to evolutions as knowledge increases.

In addition to evidence supporting a causal relationship between different events, authors are also encouraged to provide quantitative understanding of the linkage, based upon 1. Response–response relationships, time scales, known modulating factors and known positive or negative feedback loops (Society for the Advancement of Adverse Outcome Pathways, 2022).

Adverse Outcome Pathways have gained increasing regulatory acceptance but still the number of OECD approved AOPs in reprotox is low and currently limited to:

- Androgen receptor agonism leading to reproductive dysfunction
- Aromatase inhibition leading to reproductive dysfunction
- Aryl hydrocarbon receptor activation leading to early life stage mortality, via increased COX-2 or VEGF
- Inhibition of thyroperoxidase and subsequent adverse neurodevelopmental outcomes
- Histone deacetylase inhibition leading to testicular atrophy

Several additional molecular initiating events and their pathways are currently under review or open for adoption, such as histone deacetylase inhibition, estrogen receptor antagonism and PPARα activation. Nevertheless, major pathophysiological pathways for which teratogenic properties are known, still lack incorporation into the various adverse outcome networks, these include but are not limited to fetal anemia, HDAC (histone deacetylase) inhibition or methemoglobinemia, all affecting tissue differentiation.

1.2 | *In silico* models

In the directive 2010/63/EU, the European Parliament defined the Three Rs principle, described first by Russell & Burch, 1959, as aim for the protection of animals used for scientific purposes in EU. To fulfill this aim, the member states should support the research on alternative methods. At that timepoint, the focus was mainly on *in vitro* methods but with the inure of REACH regulation in 2007 the *in silico* methods became more important due to the huge amount of additional animal tests requested. For this purpose, five OECD principles were published, which have to be fulfilled by regulatory used QSARs: (a) defined endpoints; (b) unambiguous algorithm; (c) defined domain of applicability; (d) appropriate measures of goodness-of-fit, robustness, and predictivity; and (e) a mechanistic interpretation (OECD, 2006).

The great advantages of *in silico* methods are the reduction of test animals and costs and high throughput compared to animal studies (Valerio, 2009). Hence, these methods are suitable for compound selection in early developmental steps or to fill existing gaps in empirical data. This makes *in silico* methods particularly attractive for reprotox, even if the prediction is difficult due to the number and complexity of the endpoints (Hewitt, Ellison, Enoch, Madden, & Cronin, 2010). The big challenges for *in silico* prediction of reprotoxicity endpoints are the complexity of ontogenesis, the combination of several endpoints with partly unknown AOPs and the limited availability of empirical reprotoxicity data (Cronin & Worth, 2008).

1.2.1 | Model types

The available *in silico* models for reprotoxicity endpoints, which were tested in this study, is mainly Structural Alerts (SAs) and rule-based models or Quantitative Structure–Activity Relationship (QSAR) models.

SAs are chemical structures, which have been linked to toxic events (Yang, Lou, Li, Liu, & Tang, 2020). These alerts could be based on human expert knowledge (rule-based models) or generated by machine learning (Venkatapathy & Wang, 2013). Also, mixtures of both methods are common. The advantages of these models are that they are easy to interpret and allow to localize the crucial structure for toxicity. Limitations are that the methods just show the presence or absence of SAs, and absent SAs are always interpreted as non-toxicant even when based solely on incompleteness of SA lists. Besides, biological pathways of toxicity are not considered (Raies & Bajic, 2016).

QSAR models are based on the assumption that molecules that have a similar chemical structure tend to produce similar toxic effects (Hansch & Fujita, 1964). The description of the chemical structure and assessment of the similarity therefore play a decisive role in the creation

of statistical models, which are created with a training data set of sample molecules with known toxicity (Valerio, 2009). The molecules can be described by molecular descriptors, which are based on the geometric, electronic, topological, constitutional, and thermodynamic properties of the molecule (Danishuddin & Khan, 2016). However, 2D fingerprints are often used to describe the chemical structure in the form of a bit vector. In the substructure keys-based fingerprints, each bit represents the presence or absence of a predefined substructure (Cereto-Massagué et al., 2015). In contrast, topological or path-based fingerprints work by analyzing all fragments of the molecule following a path up to a certain number of bonds and then hashing each of those paths to create the fingerprint. Circular fingerprints are also hashed topological fingerprints, but they do not describe the path but the area around each atom up to a certain radius (examples for each type of fingerprint with description can be found in Table S5). Since the descriptors and the various molecular fingerprints differ greatly in their description of the molecules, this has a great influence on the functionality of the QSAR model. In order to combine the advantages of the various methods, combinations of several descriptors and a fingerprint are often used for building a QSAR model.

An alternative approach to define similarity is the use of compound class specific substructures or toxicophores (SMARTS), which can be combined with structural alerts or fingerprint techniques. This is a particular powerful approach for read across as it captures compound class intrinsic information (Enoch et al., 2022).

Significant efforts have also been invested to use bioactivity data, such as Toxcast or PubChem bioactivity data as an alternative type of descriptor. Such an affinity fingerprint is the vector consisting of compounds affinity or potency against a reference panel of proteins targets (Škuta et al., 2020). In a similar approach also effects from subchronic or chronic studies can be used. These approaches however are generally limited to marketed compounds or face the problem that bioactivity databases are proprietary information, for example, from pharmaceutical companies.

The characteristics of the models used in the study are briefly described in Table 1. A detailed description is given in Section 2.2.

1.2.2 | Importance of molecular similarity

When using prediction models, the determination of molecular similarity is of enormous importance. On one hand, this is used in QSAR models to predict toxicity and, on the other hand, it can be used for all model types to determine the applicability domain (AD). The AD is the structure space on which the training set of the model was developed, to which it is applicable to make predictions for new compounds and therefore a good benchmark if the prediction is reliable.

The choice of the description of the molecule thus has a major influence on both the prediction and its evaluation. Mellor et al. showed that the notion of fingerprint-derived similarity varies widely between data sets and structure types (Mellor et al., 2019). In particular, the subtle differences between very similar structures can often be overlooked, resulting in the same numerical similarity for such compounds. The descriptors and fingerprints for a model should therefore be selected with great care and the prediction of existing models should be critically examined by experts.

1.2.3 | Problems of prediction models for reprotoxicity

Especially for the prediction of reprotoxicity, the currently available *in silico* models have some weaknesses (Cronin & Worth, 2008). These are discussed in the following list:

- Many models only differentiate between toxic for reproduction or not. Since there can be many different MOAs with different conspicuous endpoints behind reprotoxicity, this information is very simplified. The use of models that only refer to *individual endpoints or parts of reprotoxicity* (e.g., female fertility) is therefore more promising.
- The current prediction models only consider *ADME* of pesticides indirectly via models trained on *in vivo* data. However, this is insufficient, considering, for example, the changes in the guidelines regarding exposure patterns, which can have an impact on ADME. There are currently no reliable models that can predict ADME during gestation. However, since this can greatly change the toxicity, ADME should at best be included in the models.
- To create a meaningful QSAR model, a *good quality database* is required, which should also be as comprehensive as possible. There is a lack of such data for reprotoxicity, especially since the type of data and their interpretation has changed significantly over the decades, for example, the changes in the guidelines (endpoints and dosage) and the interpretation of maternal toxicity.
- *Chemical similarity* is usually used to create QSAR models. Alternatively, or additionally, information about the compound class, biological activity or

TABLE 1 Short description of all tested models for the prediction of reprotoxicity

Platform	Model	Functional principle	Database	References
VEGA	Developmental Toxicity model (CAESAR, v.2.1.7)	• QSAR classification model • 13 EPA descriptors were used to describe molecular properties • Classifier: Random Forest	• 292 chemical compounds (mainly drug data) • 201 developmental toxicants/91 nondevelopmental toxicants • Developmental toxicity was defined by FDA categories: ○ A or B → non-toxicant ○ C, D or X → toxicant	Cassano & Benfenati (2010), Cassano et al. (2010)
	Developmental/Reproductive Toxicity library (PG, v.1.1.0)	• Empirically based decision tree • Expert rule based structural features • Molecules could be classified into 25 different categories with known DART • No categories for nontoxic chemicals • Detailed description of the categories could be found in appendix II of Wu et al. (2013)	• Decision tree is based on a data set of 716 chemicals (664 toxic, 16 non-toxic) • Detailed information about the chemicals and the references, on the basis of which they were classified, could be found in appendix I of Wu et al. (2013)	Benfenati (2020), Wu et al. (2013)
OECD (Q)SAR Toolbox	Expert-based DART scheme	• Is also based on the categories from Wu et al. (2013) like the PG model • Further development of the categories	• Same data base as PG model	OECD (Q)SAR Toolbox (2020), Wu et al. (2013)
Leadscope model applier	Repro Female Rat (RFR) v2	• Statistical based model (QSAR) • Three QSAR models were built with a balance of positive and negative compounds → prediction is the average of all three model results • Negative and positive features were identified • The predicted positive probability is based on individual contributions from the model features • Threshold in predicted positive probability is used to assign a positive or negative prediction	• Includes adverse effects to female reproductive organs (cervix, fallopian tube, ovary, uterus, and vagina) and fertility • Based on ICSAS database described by Matthews, Kruhlak, Cimino, Benz, and Contrera (2006a) • 894 training compounds (14.8% positives)	Leadscope (2021), Matthews et al. (2006a), Matthews, Kruhlak, Cimino, Benz, and Contrera (2006b)
	Repro Male Rat (RMR) v2		• Includes adverse effects to male reproductive organs (Cowper's gland, epididymis, prostate, seminal vesicles, and testes) and fertility • Based on ICSAS database described by Matthews et al. (2006a) • 714 training compounds (30.07% positives)	

TABLE 1 (Continued)

Platform	Model	Functional principle	Database	References
CASE Ultra	Foetal Dysmorphogenesis (FDYSM) Rabbit FDYSM Rat Female fertility (FFERT) Rat Male fertility (MFERT) Rat	• Statistical based SA model • Local QSAR for each alert with physicochemical descriptors • Outcome of a SAR prediction is given as the probability of being reprotoxic on a scale of 0 to 1 • Specific classification threshold for each model • Activating and deactivating alerts were detected	• Pre-processing of data and endpoints are described by Matthews et al. (2006a) • 128 active/129 inactive • 436 active/457 inactive • 113 active/113 inactive • 180 active/180 inactive	Chakravarti, Saiakhov, and Klopman (2012), Cioffi (2019a, 2019b, 2019c, 2019d), Matthews et al. (2006a), Matthews, Kruhlak, Daniel Benz, and Contrera (2007), Matthews, Kruhlak, Daniel Benz, Ivanov, et al. (2007)

SMARTS, for example, could also be used. For example, the HPPD inhibitor group of herbicides has a significant structural heterogenicity, but all are increasing tyrosine in the rat leading to respective tyrosine mediated toxicity.

- For the creation of predictive expert based structural alert models, knowledge about the *AOPs* is crucial in order to be able to correctly name the relevant structural features. Since there is still a knowledge gap for reprotox and only four AOPs have so far been recognized by the regulatory authorities, these models tend to lead to incorrect predictions.

1.3 | Testing the performance of prediction models for reprotoxicity

As discussed in the previous section, there are several challenges in predicting reprotoxicity using *in silico* models. Nevertheless, there are some commercial or freely available prediction models, which are tested in the following case studies with regard to their performance in predicting pesticides.

2 | MATERIALS AND METHODS

2.1 | Pesticide data base

To test the models with regard to their suitability for predicting reprotoxicity in pesticides, a database was created with 315 pesticides that were or are approved in the EU (see Tables S1 and S2 for pesticide DB). Five of these pesticides appeared in two versions each, which differed only in terms of stereoisomerism (cypermethrin, dimethenamid, cyhalothrin, napropamide, benalaxyl). Since most of the models to be tested do not differentiate between stereoisomers, only the 2D structures were considered in the evaluation (except for the OECD (Q)SAR Toolbox), which led to a database of 310 pesticides. In the database, the molecular structures were described by SMILES code and the InChIKeys. The reprotoxicity was assessed based on the ECHA classification according to CLP. Figure 1a shows the distribution of reprotoxicity due to ECHA classification. Notably, 256 pesticides were not classified as reprotoxicant. Notably, 17 were classified in Repr. Cat. 1B and 34 in Repr. Cat. 2. The relatively low number of potentially reprotoxic pesticides is explained by the fact that reprotoxicity is usually an exclusion criterion in the EU for the approval of a plant protection product. For the evaluation of the CAESAR model, the developmental toxicity was also selectively analyzed based on the hazard statements. Of the 51 pesticides

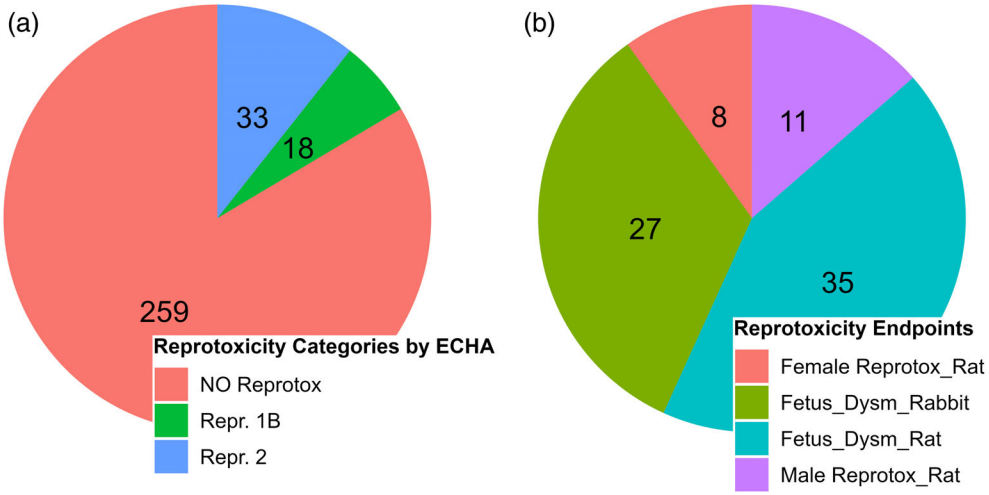


FIGURE 1 Pie charts of the distribution of (a) reprotoxicity categories by ECHA and (b) selected reprotoxicity endpoints within the pesticide DB. Pesticides classified as "NO Reprotox" do not have a Repr. 1A/B or 2 category classification but may have a classification for any other toxicity. The various reprotoxicity endpoints in chart B are based on the definition by Matthews et al. and were classified based on the studies relevant to the classification by ECHA (Matthews, Kruhlak, Daniel Benz, & Contrera, 2007)

classified as reprotoxic, only 5 are not developmentally toxic. Leadscope and CASE Ultra differentiate in their models between different endpoints of reprotoxicity: Female Reprotox_Rat, Male Reprotox_Rat, Fetus_Dysmorphogenesis_Rat, and Fetus_Dysmorphogenesis_Rabbit, which were defined by Matthews, Kruhlak, Daniel Benz, & Contrera (2007). The endpoints Female and Male_Reprotox_Rat include effects on the respective reproductive organs and fertility specific to the rat. Fetus_Dysmorphogenesis includes structural effects on fetal organs and tissues separated by species. Based on these definitions, the study results described in the available documents by ECHA/EFSA (RAC Opinion or Conclusion regarding the peer review of the pesticide risk assessment) were analyzed and corresponding columns added to the pesticide database to have comparable data. The number of pesticides per endpoint can be seen in Figure 1b. It is very important to mention, that the as reprotoxicants classified pesticides could show toxicity in one or many of these sections but also in none.

In addition, the categorization of the pesticides due to the different Resistance Action Committees (HRAC, FRAC and IRAC) and the BCPC's Compendium of Pesticide Common Names has been added, if available. The most common pesticide types in the database are fungicides, herbicides, and insecticides, a list of all types can be found in Table S3. Besides, all pesticides were classified based on their chemical structure (Chemical Group column). When categorizing according to these chemical groups, triazoles, sulfonylureas, carbamates, and organothiophosphates were the most common. Table 2 shows the 12 chemical groups with the most pesticides and the associated MOA according to the RAC poster, which applies to most of the categorized pesticides.

2.2 | Used prediction models

In the following, a selection of commercial and freely available models for DART endpoints, which were used in the case studies (see Section 3), are introduced: the open source *in silico* tools *OECD (Q)SAR Toolbox* (v4.4.1, developed by Laboratory of Mathematical Chemistry (LMC), Bulgaria, in collaboration with the Organization for Economic Co-operation and Development (OECD) and the European Chemicals Agency (ECHA), the *VEGA In Silico Platform* (v.1.1.5-b48, developed by Istituto di Ricerche Farmacologiche Mario Negri [Laboratory of Environmental Chemistry and Toxicology] and Kode srl), the commercial software packages *Leadscope Model Applier* (v3.0.2-4, developed by Instem), and *CASE Ultra* (v1.8.0.0, developed by MultiCASE Inc.). The predictions of all models and pesticides can be seen in Table S6.

2.2.1 | VEGA: Developmental Toxicity model (CAESAR, v.2.1.7)

The Developmental Toxicity CAESAR (Computer-Assisted Evaluation of industrial chemical Substances According to Regulations) model is a QSAR classification model based on a random forest method implemented using WEKA open-source libraries designed by Cassano et al., 2010. The underlying data set contains

TABLE 2 The most common chemical groups within the pesticide DB with corresponding mode of actions by the RAC-posters

Group	Pesticide type	Mode of action based on IRAC/FRAC/HRAC	#
Triazole	Fungicide	G1: Inhibition of sterol biosynthesis in membranes via C14-demethylase (19/21)	21
Sulfonylurea	Herbicide	2: Inhibition of acetolactate synthase (14/14)	14
Carbamate	Insecticide	1A: Acetylcholine esterase inhibitor (9/13)	13
Organothiophosphate	Insecticide	1B: Acetylcholine esterase inhibitor (10/10)	10
Pyrethroid	Insecticide	3A: Sodium channel modulator (9/9)	9
Aryloxphenoxypropionate (FOPs)	Herbicide	1: Inhibition of acetyl CoA carboxylase (7/7)	7
Phenoxycarboxylate	Herbicide	4: Auxin mimics (7/7)	7
Pyrazolecarboxamide	Fungicide	C2: Inhibition of succinate-dehydrogenase (7/7)	7
Strobilurin	Fungicide	C3: Inhibition of cytochrome bc1 (ubiquinol oxidase) at Qo site (cyt b gene) (7/7)	7
Phenylurea	Herbicide	5: Inhibition of photosynthesis at PSII—serine 264 binders (5/7)	7
Chloroacetamide	Herbicide	15: Inhibition of very long-chain fatty acid synthesis (6/6)	6
Dinitroaniline	Herbicide	3: Inhibition of microtubule assembly (5/6)	6

Note: The numbers in brackets indicate how many of the categorized pesticides can be assigned to the named mode of action. A list of all chemical groups can be found in Table S4

Abbreviations: FRAC, Fungicide Resistance Action Committee; HRAC, Herbicide Resistance Action Committee; IRAC, Insecticide Resistance Action Committee.

292 compounds of different classes (extracted from Arena, Sussman, Mazumdar, Yu, & Macina, 2004) whose developmental toxicity was classified according to the FDA criteria and then was subdivided in two classes: nondevelopmental toxicant (N) (FDA Cat. A and B) and developmental toxicant (D) (FDA Cat. C, D, and X). Notably, 91 compounds were classified as non-developmental and 201 as developmental toxicants. A set 13 descriptors was used for the description of the compounds calculated using Toxicity Estimation Software Tool (T.E.S.T.) (Cassano et al., 2010). The applicability of the CAESAR model is limited to organic substances with the usual elements. Predicting the toxicity of salts is only possible if they were converted into the neutralized form.

In addition to predicting toxicity using the CAESAR model, the VEGA platform itself provides an analysis of the AD. The AD is the structure space on which the training set of the model was developed and to which it is applicable to make predictions for new compounds (Hanser, Barber, Guesné, Marchaland, & Werner, 2019). To analyze the AD, the VEGA algorithm first determined the six most similar compounds within the training/test set of the model (Cassano & Benfenati, 2010). Chemical similarity was calculated by combining fingerprints with non-binary structural keys based on constitutional molecular descriptors (Floris et al., 2014). Important is that this similarity calculation is completely independent from the CAESAR model itself. Then, the two most similar compounds were used to determine the AD index (ADI), which considers also other indices besides similarity. The ADI has values from 0 (worst case) to 1 (best case) and is the basis for the reliability classes good, moderate, and low. Since the training set is not very large, the provided information about the similar compounds and the ADI are very useful to evaluate the prediction.

The validation statistic states the sensitivity as 95% and the specificity as 59%. This is sensible since the models have been developed with the aim to minimize false negatives in order to make the CAESAR model usable for REACH (Cassano et al., 2010). This tendency has to be considered when analyzing the predictions.

2.2.2 | OECD (Q)SAR Toolbox

The expert-based developmental and reproductive toxicity (DART) scheme (v.1.4, developed by Procter & Gamble and LMC) is based on a decision tree for identifying chemicals as developmental and/or reproductive toxicants presented in Wu et al., 2013. This decision tree was designed based on the combination of known modes of action (MOA) and associated structural features, as well as an empirical association of structural fragments within DART chemicals when MOA information was not available. According to Wu et al., 2013, the decision tree was not originally intended to be used as a standalone predictive tool, but as part of a screening system to identify potentially reproductively toxic chemicals and as part of

weight-of-evidence-based structure–activity relationship (SAR) decisions. This conflicts with use by the expert-based DART scheme of the OECD (Q)SAR Toolbox.

The stereochemistry of the test substances is relevant to the prediction in this model, accepting the nine categories (2, 3, 4, 5, 6, 7, 14, 16, and 18) where stereoisomerism is ignored. Besides, the applicability of the model is limited to organic substances. The profiler's database comprises 716 chemicals (664 positive, 16 negative and 36 with insufficient data) that were investigated for their DART potential (OECD (Q)SAR Toolbox, 2020; Wu et al., 2013). It includes 25 different categories and 129 sub-categories, based on defined receptor binding and chemical properties and, if known, their MOA. It should be noted that the tool is not intended as a stand-alone system to support regulatory decision-processes (OECD (Q)SAR Toolbox, 2020).

2.2.3 | VEGA: Developmental/Reproductive Toxicity library (PG, v.1.1.0)

The PG (Procter&Gamble) model is an empirically based decision tree designed by Wu et al., 2013 (see Section 2.2.2) and, therefore, is very similar to the DART model of the OECD (Q)SAR Toolbox. However, the PG model is available at the VEGA platform, which automatically calculates the most similar compounds of the training set, an applicability domain index (ADI) and based on this, indicates a reliability (Benfenati, 2020). This additional information is very helpful for assessing the prediction, for example, to check the classification in a certain category based on similar compounds. In this model, pesticides are predicted to be non-toxic, if their core structural features fall outside of the chemical domains covered by the DART decision tree. It is important to realize that the PG model, by design, is incapable of predicting non-reproductively toxic substances, as there are no such categories. The correct description would be that there is no known DART precedent, which does not automatically imply the absence of DART endpoint effects (Wu et al., 2013). As with the CAESAR model, the sensitivity here at 0.89 is significantly greater than the specificity at 0.44 (Benfenati, 2020).

2.2.4 | Leadscope Model Applier

The statistical models used in the Reproductive Toxicity Suite, Repro Female Rat (RFR) v2 and Repro Male Rat (RMR) v2, are intended to be used in screening, prioritization and can be used in a weight of evidence approach particularly for designing studies and interpretation of findings, which may be used in regulatory contexts (Leadscope, 2021). These models were developed under a Research Collaboration Agreement (RCA) with the United States Food and Drug Administration (FDA) (Leadscope, 2021). The training set of the RFR model consists of 894 structures and that of the RMR model consists of 714 (Leadscope, 2021), which were obtained from the Informatics and Computational Safety Analysis Staff (ICSAS) database described in Division of Applied Regulatory Science (DARS) publications of the FDA (Matthews et al., 2006a, 2006b). The training set of the Leadscope RFR model includes adverse effects on the female reproductive system and fertility, while it does not include effects on the fetus, gestation, or lactation. Reprotoxicity in the RMR model comprises adverse effects on the reproductive system and fertility in male rats (Matthews et al., 2006a). The RFR model comprises 14.08% positives in its training set, while the RMR model includes 30.07%. Because of the unbalanced nature of the training sets, each model combines the results of three sub-models with balanced sets as average model (Leadscope development team, personal communication).

The structural features identified by the models are either positively or negatively correlated with activity. Such features are highlighted in the structure to facilitate a rapid review of features which are associated with activity and to assess the coverage of the structural elements by the models. This information is provided in the prediction report. The following eight property descriptors are used in the RFR and RMR models: A Log*p*, polar surface area, hydrogen bond acceptors, rotatable bonds, parent molecular weight, hydrogen bond donors, parent atom count, and Lipinski score (Leadscope development team, personal communication).

The Leadscope software uses the following parameters to manage the AD of the models: in addition to all property descriptors, at least one structural feature and at least one chemical in the training set with at least 30% global similarity to the test chemical is required to generate predictions (Leadscope, 2021). The similarity score is based on Leadscope's 27,000 sub-structural features and hence will be lower than similarity scores that use smaller feature sets.

2.2.5 | MultiCASE Software

CASE Ultra is a commercial tool by MultiCASE, which provides classification models for different reproductive and developmental toxicity endpoints based on *in vivo* data for mouse, rat, or rabbit from FDA as part of Research Cooperation Agreement (RCA). Four of these

endpoint models (see Table 3) were selected for evaluation of their predictive power for pesticides. The definition of these endpoint models was given by Matthews et al., 2007 and the pre-processing of data before modeling was published in Matthews et al., 2006a. The models were based on different data sets, but all were statistical based SA models, which were built by collecting positive or deactivating alerts from the training data set that are related to the toxicity being modeled (Chakravarti et al., 2012). In addition, a local QSAR was built for each alert with physicochemical descriptors. The outcome of the prediction was given as the probability of being reprotoxic on a scale of 0 to 1 and by use of the classification threshold (CT) (specific for each model) the prediction was done. A probability between 0 and CT−0.1 leads to a negative or out of domain classification. When the probability is between CT−0.1 and CT+0.1 the substance is classified as inconclusive and above CT+0.1 as positive. The AD of the model is defined by a fragment based chemical space defined by the training set chemicals (Cioffi, 2019a, 2019b, 2019c, 2019d). The AD is assessed by checking for 3-atom fragments that are not present in the trainings set. Due to the limits of applicability inorganic compounds, mixtures and large biomolecules are in principle not covered by the AD. The prediction report provided by CASE Ultra contains detailed information about the alerts and structural analogs etc., which are of great importance when assessing the prediction.

2.3 | Evaluation of predictions

Analysis was conducted in KNIME (version 4.3.2) (Berthold et al., 2008) and R (version 4.0.2) (R Core Team, 2019) and figures were produced using the R package ggplot2 (Wickham, 2016). All shown chemical structures were copied from PubChem or from the respective model reports. For the assessment, the predicted toxicity of the PG and the QSAR Toolbox model was compared with the classification by the ECHA. In the case of the CAESAR model, predictions were compared to developmental determined by ECHA. For the Leadscope and CASE Ultra models, the results of the animal experiments in rats and rabbits on which the ECHA classification is based were used. If the pesticide was predicted as nontoxic the evaluation could be True Negative (TN) or False Negative (FN) and if the prediction was toxic the possible evaluations were True Positive (TP) or False Positive (FP). In the case that no reliable prediction could be made the evaluation is UNKNOWN (see Table 4).

Besides the typical values of an error matrix (TN, FN, TP, FP), also the sensitivity (SEN), specificity (SPC), accuracy (ACC), and balanced accuracy (BA) were

TABLE 4 List of possible predictions of all models and the resulting evaluations

Model	Prediction	Evaluation
VEGA_CAESAR	NON-toxicant (experimental value, good/moderate reliability)	TN, FN
	Toxicant (experimental value, good/moderate reliability)	TP, FP
	NON-toxicant/toxicant (low reliability)	UNKNOWN
VEGA_PG	NON-toxicant	TN, FN
	Toxicant	TP, FP
OECD (Q)SAR Toolbox (OQTB)	Not known precedent reproductive and developmental toxic potential	TN, FN
	Known precedent reproductive and developmental toxic potential	TP, FP
	Not covered by current version of the decision tree	UNKNOWN
Leadscope (LS)	Negative/Negative_EV	TN, FN
	Positive/Positive_EV	TP, FP
	Missing descriptors/not in domain	UNKNOWN
CASE Ultra (CU)	Negative/known negative	TN, FN
	Positive/known positive	TP, FP
	Inconclusive/out of domain	UNKNOWN

Abbreviations: FN, false negative; FP, false positive; TN, true negative; TP, true positive.

TABLE 3 The properties of the evaluated CASE Ultra models

Model	Description	Species	# Active/inactive	# Descriptors	Classification threshold
FDYSM	Fetal Dysmorphogenesis	Rabbit	128/129	19	0.5
		Rat	436/457	111	0.45
FFRET	Female fertility	Rat	113/113	47	0.55
MFRET	Male fertility	Rat	180/180	47	0.5

determined (for definitions see Table 5). For models that predict toxicity then sensitivity is a more critical value than the specificity, because of safety reasons FP are much more tolerable than FN. In this study, the pesticide DB and also most of the training sets of the models were unbalanced data sets; therefore, the BA is calculated besides the more common ACC.

3 | RESULTS AND DISCUSSION

3.1 | CAESAR model of VEGA

The statistical developmental toxicity QSAR model CAESAR is based on a data set of 292 compounds whereby the majority of the compounds was classified as "Toxicant" (69%). The model is available via the VEGA platform, which provides an assessment of the reliability in addition to the prediction. This reliability relates to whether the connection is inside or outside the model's AD.

3.1.1 | Evaluation

Table 6 illustrates the reliability distribution within the tested pesticide database. The vast majority of pesticides were outside the AD of the model (77%). This shows that the model cannot provide a meaningful prediction for most pesticides.

The predictions were evaluated by comparing them with the GHS classification of ECHA referring to developmental toxicity (see M&M). Pesticides whose prediction was classified as unreliable (Out of AD) were classified as UNKNOWN for the evaluation. This resulted in a large number of false positives, especially for the pesticides, the prediction of which was classified as good (see Table 7). This observation agrees with the results of the published validation of the model (Cassano & Benfenati, 2010), which also show a high proportion of false positives and thus a low specificity. This is due to the high overhang of toxic compounds in the training data set of the CAESAR model and is reinforced by the opposite distribution in the pesticide database. The proportion of false negatives, on the other hand, is very low, which leads to a sensitivity of 0.89.

3.1.2 | Example

The following example shows in detail why false predictions are made despite good reliability (ADI > 0.8): Napropamide is an herbicide that belongs to the chemical group of acetamides. According to the GHS classification, napropamide is not toxic to development, but was classified as developmental toxicant by the CAESAR model, therefore as a false positive prediction. The reliability was given as good (ADI = 0.918), which means that napropamide was within the AD of the model. This classification is based on the two most similar compounds in the training data set of the model and their classification, which can be viewed in the report. The two most similar substances were Phenyltoloxamine and Naproxen with a similarity score of 0.855 and 0.83 (see Table 8). According to the model, connections with similarity scores above 0.75 are to be regarded as sufficiently similar. This seems questionable when comparing the chemical structure of napropamide with phenyltoloxamine and naproxen. The

TABLE 5 The formulas for calculating the typical parameters to evaluate prediction models

Value	Name	Definition
SEN	Sensitivity	$TP/TP+FN$
SPC	Specificity	$TN/TN+FP$
ACC	Accuracy	$TP+TN/TP+TN+FP+FN$
BA	Balanced accuracy	$SEN+SPC/2$

Abbreviations: FN, false negative; FP, false positive; TN, true negative; TP, true positive.

TABLE 6 The distribution of reliability of the developmental toxicity prediction of 310 pesticides using the CAESAR model provided by VEGA

Reliability	Applicability domain	#	# [%]
Experimental value	The predicted compound *could be out* of the Applicability Domain of the model	1	0.32
Good reliability	The predicted compound *is into* the Applicability Domain of the model	28	9.03
Moderate reliability	The predicted compound *could be out* of the Applicability Domain of the model	40	12.90
Low reliability	The predicted compound *is outside* the Applicability Domain of the model	241	77.74

TABLE 7 The results of evaluation of the CAESAR model via typical parameters

	# FN	# FP	# TN	# TP	# UNKNOWN	SEN	SPC	BA	ACC
ALL	2	42	8	17	241	0.89	0.16	0.53	0.36
Experimental value	0	0	1	0	-	-	1.00	-	1.00
Good reliability	1	23	1	3	-	0.75	0.04	0.40	0.14
Moderate reliability	1	19	6	14	-	0.93	0.24	0.59	0.50

Note: In addition to the evaluation for all pesticides, the following lines contain the evaluation related to the prediction reliability.
Abbreviations: ACC, accuracy; BA, balanced accuracy; FN, false negative; FP, false positive; SEN, sensitivity; SPC, specificity; TN, true negative; TP, true positive.

main structures of napropamide methoxynaphthalene and acetamide were not mirrored. In addition, phenyltoloxamine was only part of the test set and therefore did not serve as the basis for the model. Overall, similarity scores should be viewed critically, since the values depend heavily on the choice of descriptors, as can be seen in Table 8. Next, the toxic classification of the similar compounds is considered. Both were classified as toxic to development due to their FDA classification, while the ECHA only classifies naproxen as developmentally toxic.

3.1.3 | Summary

In summary, many wrong predictions despite good reliability were made because of an insufficiently similarity of the "most similar compounds" as well as different data sources for the assessment of toxicity. Therefore, the similarity of the compounds and the data sources should always be checked when assessing the prediction.

3.2 | PG model of VEGA

The PG model for the prediction of DART is available via the VEGA platform. It is a rule-based model where the classification takes place via a decision tree. The compounds are categorized into 25 different chemical categories including several subgroups. It is important that the established rules are only suitable for the detection of DART, but that there are no rules that describe non-DART structures. Therefore, the consideration of reliability only makes sense for categorized and thus classified as toxic pesticides, since all others should not be within the AD of the model by definition (see Table S8).

3.2.1 | Evaluation

When analyzing the results, a large number of pesticides (39, 12.5% of all pesticides, see Table S8) were labeled as experimental value, which means that they can also be found in the training data set of the model. Of these, 59% were false positive, which indicates a different interpretation of DART in the data set of the PG model and by the ECHA (see Table 9).

In predicting the DART of the pesticide database, 216 pesticides were not categorized and thus classified as non-toxic. All other categorized were divided into 14 categories, with categories 1 (inorganics and derivatives metals, metallic derivatives, organophosphorus and organosiloxane compounds), 8 (aromatic compounds with alkyl, multi-halogen and nitro groups), and 13 (imidazole, nitro imidazoles derivatives, nitro-furfurylideneamino and triazole derivatives) being the most common (see Figure 2a and Table S9).

Overall, the model classified two-thirds of the pesticides that are toxic to reproduction as non-toxic (see Table 9). These belonged to different chemical groups and were either not assigned to the "right" category or there was no suitable category. The proportion of false positives was 25%. Figure 3 shows the distribution of FP and TP per category. In all categories, the number of FPs was higher than the number of TPs except for category 13, which includes triazole and imidazole.

3.2.2 | Example

In the following, the evaluation of the prediction is shown on the basis of the PDF report provided by VEGA using the example of 2,4-dichlorophenoxyacetic acid (2,4-D) (see Table 10). 2,4-D is a phenoxy herbicide belonging to the auxins group. It was classified as toxicant due to experimental value, which was different to the ECHA classification. 2,4-D was categorized into category 8c (aromatic compounds with alkyl, multi-halogen and nitro groups, examples: para-dichlorbenzene, 1,2,4-trichlorobenzene) based on the dichlorobenzene sub-structure (matching rule/virtual compound, see Table 10). The phenoxy acetic acid part was not taken into account, which leads to a misleading categorization. All six most similar compounds

TABLE 8 Example for a false positive CAESAR prediction despite good reliability

	Name	Structure	Developmental toxicant?	CAESAR prediction	Similarity by VEGA	Average Tanimoto similarity coefficient			
						Pub chem	RDKit	Morgan	Feat Morgan
Tested pesticide	Napropamide		NO (ECHA)	Toxicant		1	1	1	1
Similar compound 1	Phenyltoloxamine		YES (CAESAR)	Toxicant	0.855	.828	.345	.208	.205
Similar compound 2	Naproxen		YES (CAESAR)	Toxicant	0.83	.671	.331	.212	.282

Note: Napropamide was the tested pesticide and phenyltoloxamine and naproxen were the most similar compounds of the data set of the CAESAR model. To show the variability of similarity depending on the selected descriptors, the similarity score provided by the VEGA platform was compared with the average Tanimoto similarity coefficient based on different fingerprints.

provided by the VEGA software were phenoxy herbicides and categorized to category 9c (alpha aryloxy substituted acetic acid, examples: 2,4,5-trichlorophenoxyacetic acid, 2,4-D Isopropyl ester), which would to be a much more suitable category also for 2,4-D. Since the prediction was false positive, although 2,4-D is part of the training data set of the PG model, the data source was of great interest. The DART toxicity of 2,4-D was described by the Reproductive respectively Developmental Toxicity Effect Codes R(T) (Changes in reproductive function/fertility only occurred at doses where there was significant toxicity on other organ systems) and D(MT) (Developmental effects only occur in the presence of maternal toxicity) and the Reregistration Eligibility Decision document by U.S. EPA was given as reference U.S. EPA, 2005a, 2005b. In this case, the problem lies in the fact that the classification is based on different study data, or the study data were interpreted differently.

3.2.3 | Summary

The PDF report of the PG model describes exactly based on which structure fragment the pesticide was classified in the respective category. The more structures of the original molecule are covered, the better. In contrast to most models, the PG model also offers a detailed description of the sources on the basis of which the compounds in the data set were classified. All of this information should be considered when assessing the prediction.

3.3 | DART scheme of OECD (Q)SAR Toolbox

The aim of the Developmental and Reproductive Toxicity (DART) scheme implemented in the OECD (Q)SAR Toolbox is to indicate that the test compound is associated with chemical structures known to have DART, or that it contains structural features that are outside the AD of the DART decision tree (OECD (Q)SAR Toolbox, 2020). The Toolbox's DART scheme is a rule-based profiler in which the classification is carried out using a decision tree (OECD (Q)SAR Toolbox, 2020), similar to the PG model. This decision tree includes 25 different chemical categories including 129 subcategories (OECD (Q)SAR Toolbox, 2020). It should be noted that the established rules are only suitable for the detection of DART, but there are no rules that describe non-DART structures. In contrast to the PG model, with the Toolbox's DART scheme there is no structure-based comparison of the predicted substance with the training data set and therefore no evaluation of the AD.

TABLE 9 The results of evaluation of the PG model via typical parameters

	# FN	# FP	# TN	# TP	SEN	SPC	BA	ACC
ALL	34	77	182	17	0.33	0.70	0.52	0.64
Experimental value	0	23	3	13	1.00	0.12	0.56	0.41
Categorized	0	77	0	17	1.00	0	0.50	0.18
Uncategorized	34	0	182	0	0	1.00	0.50	0.84

Note: In addition to the evaluation for all pesticides, the following lines differentiate between experimental value and categorized or uncategorized pesticides.
Abbreviations: ACC, accuracy; BA, balanced accuracy; FN, false negative; FP, false positive; SEN, sensitivity; SPC, specificity; TN, true negative; TP, true positive.

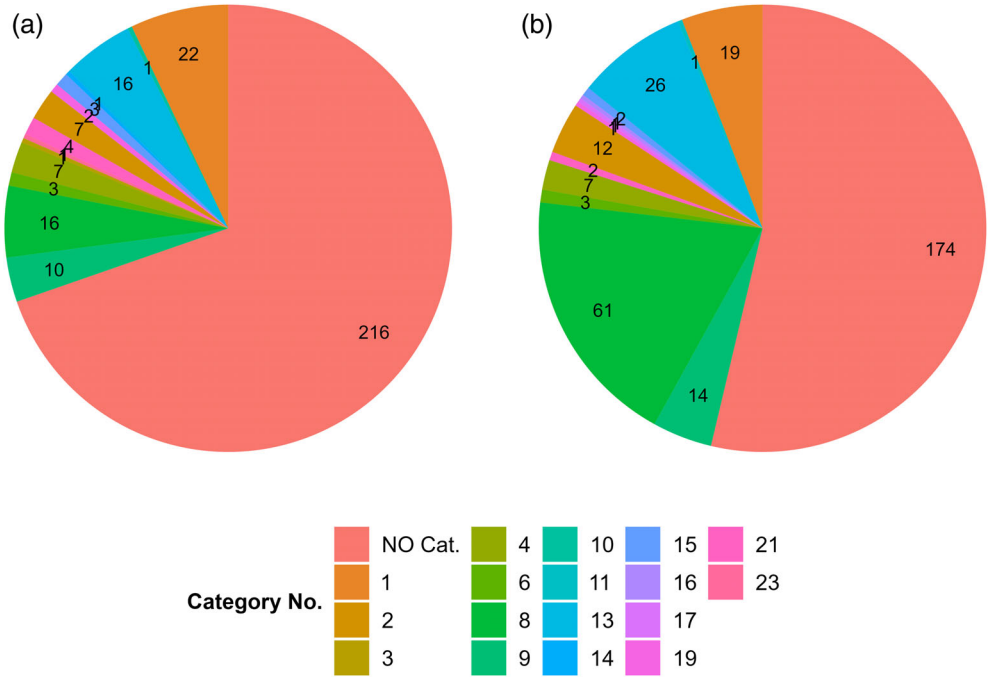


FIGURE 2 The pie charts show the distribution of pesticides in the chemical categories defined by Cassano et al. (2010) predicted by the PG model (a) or DART scheme of the OECD (Q)SAR Toolbox (b). The structural description of the categories can be found in Table S7

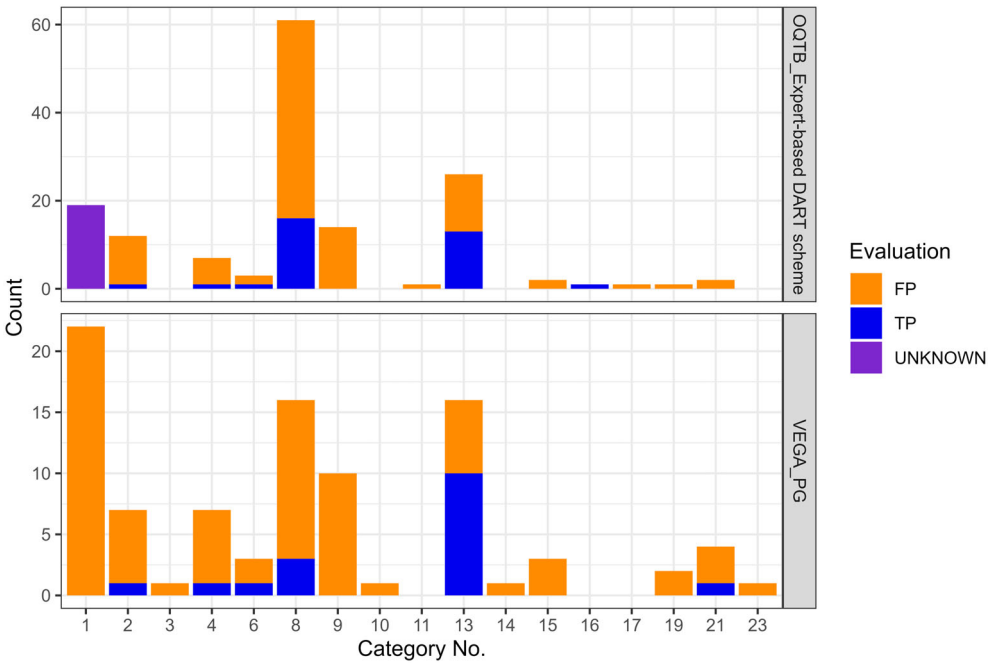
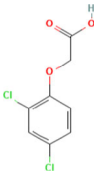
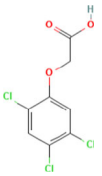
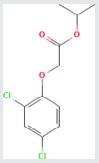


FIGURE 3 The bar plots show the evaluation of the predictions divided by the predicted categories for the PG and DART model by OECD (Q)SAR Toolbox. The aim of the depiction is to analyze whether the prediction for some categories is more reliable than for others. FP, false positive; TP, true positive

TABLE 10 Structure of the tested pesticide 2,4-D, the matching rule/virtual compound and the two most similar compounds, as well as their predicted categories by the PG model

	Name	Structure	Predicted category
Tested pesticide	2,4-D		8c
Matching rule/virtual compound	-		-
Similar compound 2	2,4,5-trichlorophenoxyacetic acid		9c
Similar compound 3	2,4-D isopropyl ester		9c

The pesticides were classified into one of the following three categories: "Not known precedent reproductive and developmental toxic potential," "Known precedent reproductive and developmental toxic potential," or "Not covered by current version of the decision tree." The latter category means that the test compound is out of AD of the DART profiler. Thus, such compounds are not classified by the DART scheme (OECD (Q)SAR Toolbox, 2020). In addition, the following subcategories were identified for the pesticides tested in this category: "Inorganic chemical," "Metal atoms were identified, Metals (1a)," or "Organophosphorus compounds (1b)." For the evaluation of the pesticide predictions, the category that was identified outside of the AD is interpreted as UNKNOWN. Further, if a substance does not match one of the structural features associated with the potential to act as a DART compound, it is classified as "Not known precedent reproductive and developmental toxic potential."

In contrast to the other models, the DART model of the OECD (Q)SAR Toolbox distinguishes between stereoisomers, which is why these were also used for the prediction when relevant. There were no differences in the prediction between the stereoisomers of the tested pesticides or in comparison with the 2D structures.

3.3.1 | Evaluation

Most pesticides (56%) were not associated with chemical structures known to have DART and were therefore identified as negative, 37% were predicted as positive, and 6% were outside of the profiler's AD. In comparison, the similar PG model implemented in the VEGA platform predicted 70% of the pesticides as negative, 30% as positive, and 0% were outside the AD of the model.

In the DART prediction of the pesticide database, a total of 136 categorized pesticides (117 DART positives and 19 compounds that were outside the AD) were divided into 13 categories (Figure 2b) and 18 subcategories. Of these 150 pesticides, 14 were categorized into two categories. The higher proportion of 174 pesticides was not assigned to any category (56%) and was therefore predicted as negative. Table S10 shows the distribution of pesticides in the respective categories and subcategories.

In the following, either all prediction results of the Toolbox's DART scheme or only uncategorized and categorized results are examined (Table 11) and compared with VEGA's PG model.

When investigating all of the pesticide predictions, the Toolbox's DART profiler predicted 43% of the pesticides that are DART positive as non-toxic compared to the ECHA GHS classification. This is a better prediction result compared to the similar PG model, which classified 66% as non-DART. This group of false negative tested pesticides similarly includes different pesticide types and chemical groups in both models. The false negatives were either not classified to the "right" category or there was no appropriate category. A sensitivity of 57% and a specificity of 63% were identified in the Toolbox's DART

TABLE 11 The results of evaluation of the DART scheme of the OECD (Q)SAR Toolbox via typical parameters

	# FN	# FP	# TN	# TP	# UNKNOWN	SEN	SPC	BA	ACC
ALL	22	88	152	29	19	57	63	60	62
Categorized	0	88	0	29	0	100	0	50	25
Uncategorized	22	0	152	0	0	0	100	50	87

Note: In addition to the evaluation for all pesticides, the following lines differentiate between categorized and uncategorized pesticides.
Abbreviations: ACC, accuracy; BA, balanced accuracy; FN, false negative; FP, false positive; SEN, sensitivity; SPC, specificity; TN, true negative; TP, true positive.

scheme, while a sensitivity of only 33% and a specificity of 70% were revealed in the PG model.

Of the 174 uncategorized pesticides identified as negative in the Toolbox were 87.4% classified as true negative and 12.6% as false negative. The PG model showed similar results, in which 216 uncategorized pesticides were identified 84.3% as true negative and 15.7% as false negative.

More pesticides were classified in the Toolbox than in the PG model. Of the 117 categorized DART positives in the Toolbox were 24.8% classified as true positive and 75.2% as false positive, while of the 94 classified DART positives in the PG model were 18.1% identified as true positive and 81.9% as false positive. The distribution of true positive and false positive predictions per chemical category in the Toolbox is presented in Figure 3. In all categories with a higher number of pesticides (i.e., ≥ 7 pesticides/category), the number of false positives was higher than that of true positives, with the exception of Category 13, which included both with the same frequency. Category 13 includes triazole and imidazole. These results are similar to those of the PG model with the exception that more true positives were recognized in Category 13. Further, as mentioned above, more pesticides were categorized using the Toolbox's DART scheme than the PG model. In particular, the number of pesticides in chemical Category 8 (above all "Toluene and small alkyl toluene derivatives (8a)" and "Polyhalogenated benzene derivatives (8c)") was much higher with the Toolbox's DART profiler than with the PG model (Figure 3).

It should be kept in mind that both models contain unbalanced training sets, with 92.7% positives, only 2.2% negatives, and 5% substances with insufficient data in their databases (OECD (Q)SAR Toolbox, 2020; Wu et al., 2013). Only DART positive structural alerts are used to categorize the test substances. This strong imbalance in the direction of DART positives in the training set may cause the high number of false positive results.

In general, the Toolbox's DART profiler has a slightly better statistical profile in terms of DART prediction compared to the VEGA's PG model. However, both DART models tend to predict a higher number of false positives and therefore show low specificity. Hence, both systems are "overcautious" and may hinder the regulatory decision-process of pesticides.

One of the model differences is that the Toolbox's DART profiler classifies all pesticides of Category 1 ("Inorganic chemical," "Metal atoms were identified, Metals (1a)," and "Organophosphorus compounds (1b)") as "Not covered by current version of the decision tree" (UNKNOWN; Figure 3), while the pesticides in Category 1 ("Inorganics and derivatives: metals, metallic derivatives, organophosphorus, and organosiloxane compounds") of the PG model are assigned as toxicants. However, when comparing the Category 1 pesticides of both models with the ECHA GHS classification, none of them were classified as DART positive. It can therefore be concluded that an incorrect classification was implemented in the PG model for Category 1 substances.

A more detailed comparison of the predictions of both models shows that of the 310 pesticides tested, 157 were not categorized by both models, 66 were assigned similarly, and only 2 pesticides (1,4-dimethylnaphthalene and 2,4-D) were classified in different categories. Further, 54 were only categorized by the DART profiler of the Toolbox and 17 only by the PG model (see Figure S1A). Of the 54 pesticides categorized only by the Toolbox's DART profiler, most were classified as Category 8 ("Toluene and small alkyl toluene derivatives (8a)": 29 and "Polyhalogenated benzene derivatives (8c)": 10) and 13 ("Triazole derivatives (13c)": 10) (Figure S2A). In contrast, of the 17 pesticides that were only categorized by the PG model, most of them were assigned to Category 1 ("Inorganics and derivatives: metals, metallic derivatives, organophosphorus and organosiloxane compounds": 4) and 8 ("Aromatic compounds with alkyl, multi-halogen, and nitro groups": 3) (see Figure S2B). In addition, the Toolbox's DART profiler assigned 14 pesticides to 2 categories, 5 (e.g., fluquinconazole) of which were only categorized by the DART profiler, for 8 (e.g., penconazole) one categorization was similar to the PG model and the other was not and for 1 (dicloran) both categories of the DART profiler were similar to the PG model (Figure S1B).

3.3.2 | Example

In a group-based case study, the classification of DART positive pesticides in the Subcategory "Toluene and small alkyl toluene derivatives (8a)" by the Toolbox's DART profiler is investigating in the following.

The structural framework of this subcategory implemented in the Toolbox is presented in Wu et al., 2013 and further developed by Procter & Gamble and LMC, Bulgaria (OECD (Q)SAR Toolbox, 2020). Only toluene and a single attached alkyl chain substituent (< 5 carbon atoms) are structural features of this category according to Wu et al., 2013 (Figure 4). The possible alkyl chain substituents can be at ortho-, para- or meta-positions. Members of the training data set (e.g., toluene, p-xylene or butyltoluene) meet these conditions (Wu et al., 2013).

It is noticeable that all 29 pesticides assigned to Subcategory 8a by the Toolbox contain, in addition to toluene, larger substitutes (> 5 carbon atoms; including N, O, Cl, F, S, or Br atoms) that are not described in the original category definition of Wu et al., 2013 (selected pesticides shown in Table 12). Therefore, the categorization in Subcategory 8a is considered wrong, since the pesticides do not belong to the chemical class of toluene and small alkyl toluene derivatives. The similar PG model, on the other hand, which is closer to the description of Wu et al. (2013), did not classify any of the pesticides in Subcategory 8a.

In conclusion, the classification of the 29 pesticides in Subcategory 8a is overall wrong or is not based on the requirements described in Wu et al., 2013. Hence, the Toolbox's DART profiler is not reliable to predict the DART potential of pesticides that contain toluene and alkyl toluene derivatives.

3.3.3 | Summary

The case study with toluene and alkyl toluene derivatives illustrates well the general problem of the QSAR prediction for pesticides using the Toolbox's DART scheme. Many false negative and false positive predictions were generated with the Toolbox, probably mainly due to incorrect classification of pesticides into different chemical categories. Therefore, when evaluating the predictions, care should be taken to ensure that the categorization of the chemical classes is correctly chosen by the Toolbox.

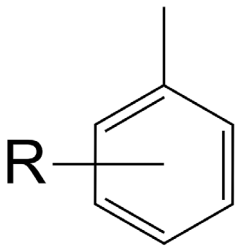


FIGURE 4 The structural scope of "Toluene and small alkyl toluene derivatives (8a)." R = H, Me, nBu, iPropyl, tBu

3.4 | Leadscope

In the present publication, prediction results from Repro Female Rat (RFR) and Repro Male Rat (RMR) statistical QSAR models of the Reproductive Toxicity Suite were analyzed. The training set of the Leadscope RFR model includes only adverse effects on the female reproductive system and fertility, while it does not include effects on the fetus, gestation, or lactation. Reprotoxicity in the RMR model only comprises adverse effects on the reproductive system and fertility in male rats (Matthews et al., 2006a). Therefore, the predictions were compared with the results of the experiments relevant for the classification according to ECHA, based on the endpoints mentioned.

Both QSAR models assess potential reprotoxicity of test substances based on a statistical weighting of structural features present in the test structures as well as whole molecule descriptors. If experimental data are available within the Reproductive Toxicity Suite, these data will be used instead of the QSAR prediction. Probability scores below the cut off value of .5 are negative and values equal to or greater than .5 are considered positive (see Figure S3).

3.4.1 | Evaluation

When analyzing the prediction results of both models (Table 13), about half of the pesticides were classified as "UNKNOWN," which comprises "Missing Descriptors" and "Not in Domain" calls (RFR: 47%, RMR: 59%) (see Table S11). A "Not in Domain" call indicates that the predicted pesticides were outside the model's AD. In the case of "missing descriptors," this is due to inorganic structures for which the whole molecule descriptors cannot be calculated. Due to this classification, 4 reprotoxicants and 143 non-reprotoxicants could not be predicted by the RFR model, and 4 reprotoxicants and 179 non-reprotoxicants were not recognized by the RMR model when the predictions were compared with the ECHA GHS classification. Thus, 50 or 36% of the reprotoxicants could not be detected for each model, as they were outside the AD.

TABLE 12 A selection of pesticides that were incorrectly classified in subcategory 8a

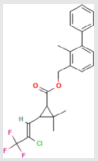
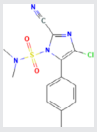
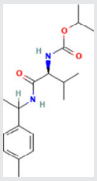
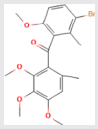
Name	1,4-dimethyl-naphthalene	Bifenthrin	Cyazofamid	Iprovalicarb	Metrafenone
CAS no.	571-58-4	82657-04-3	120116-88-3	140923-17-7	220899-03-6
Structure					

TABLE 13 The results of evaluation of the two Leadscope models Repro Female Rat (RFR) and Repro Male Rat (RMR) via typical parameters

Model	# FN	# FP	# TN	# TP	# UNKNOWN	SEN	SPC	BA	ACC
RFR	3	3	156	1	147	0.25	0.98	0.62	0.96
RMR	4	26	94	3	183	0.43	0.78	0.61	0.76

Abbreviations: ACC, accuracy; BA, balanced accuracy; FN, false negative; FP, false positive; SEN, sensitivity; SPC, specificity; TN, true negative; TP, true positive.

The Leadscope RFR and RMR models contain experimental data of the Informatics and Computational Safety Analysis Staff (ICSAS) database (Matthews et al., 2006a) with data records from FDA segment I (reprotoxicity in male and female rats) studies. The data were obtained from publicly available sources (e.g., Shepard's Catalog of Teratogenic Agents, TERIS, REPROTOX, and RTECS), as well as studies reported in drug labeling (Matthews et al., 2006a).

When comparing the prediction of pesticides, which are not only available in the pesticide database, but also in the training set, it is noticeable that these pesticides with experimental data are also sometimes "incorrectly" predicted (see Table S12). This is due to differences in the data from the Reproductive Toxicity Suite and the ECHA data set. The reason for these differences can be, for example, different *in vivo* studies on which the assessment is based or different evaluation of reprotoxic effects as adverse or not adverse.

Both models show a low sensitivity (RFR: 0.25, RFM: 0.43) in predicting the specific reprotoxicity endpoints of pesticides, while the specificity is high (RFR: 0.96, RFM: 0.76) (Table 13). If the statistical profile of the pesticide predictions is compared with those of the organic chemicals in the Leadscope manual, the sensitivity for the organic chemicals is much higher in both models (RFR: 61%, RFM: 85%), while the specificity is comparable (RFR: 95%, RFM: 73%) (Leadscope, 2021). The low sensitivity of both Leadscope models confirms that they should not be used in isolation for the regulatory evaluation of pesticides and additional lines of evidence such as through an expert review of model features and potentially reactive features, a consensus approach using predictions from other models in the Reproductive Toxicity Suite and/or experimental findings are needed.

As mentioned above, structural features and property descriptors are used to determine a probability score that drives the prediction. The distribution of probability scores per prediction can be seen in Figure S3. It is crucial for the assessment of the prediction, that the positive or negative statement is based on a threshold value (0.5) and a connection with the stability of the prediction and the absolute value of the probability cannot be considered without other information.

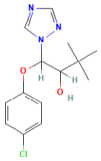
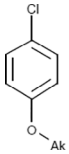
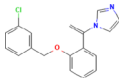
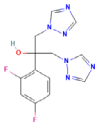
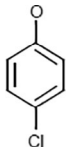
3.4.2 | Examples

In a group-based case study, the prediction results of the structurally diverse conazole fungicides (imidazoles and triazoles) from the two selected RFR and RMR models were analyzed (see Table S13). Conazoles, a class of azole-based fungicides, are widely used as pesticides, but also as human pharmaceuticals to treat mycoses (Kjærstad, Taxvig, Nellemann, Vinggaard, & Andersen, 2010; Zarn, Brüschweiler, & Schlatter, 2003) during pregnancy (King, Rogers, Cleary, & Chapman, 1998; Mogensen et al., 2017).

Of the 24 conazole fungicides included in the EFSA conclusions, two substances (epoxiconazole and triadimenol) are reprotoxic in female rats and one (triadimenol) in male rats according to their ECHA GHS classification. However, only one of the tested conazoles (i.e., epoxiconazole) was correctly predicted as reprotoxic by the RFR model, while the RMR model identified the substance false positive. The other reprotoxic pesticide, triadimenol, was either classified as false negative in the female model or outside the AD in the male model.

The negative prediction of triadimenol by the RFR model, due to the detected structural feature contribution of benzene, 1-alkoxy, 4-chloro (Table 14) and the

TABLE 14 Detected structural features and selected training set analogs of triadimenol, which is reprotoxic in female and male rats

Predicted pesticide	Model	Evaluation	Detected structural features	Selected relevant analog structures	
Triadimanol CAS no. 55219-65-3	RFR v2	FN	Benzene, 1-alkoxy-, 4-chloro-	*Croconazole* *Positve* for RFR	*Fluconazole* *Negative* for RFR
					
	RMR v2	UNKNOWN	Chlorophenol-	No analog structures reported.	
					

property descriptors, which resulted in a probability score of .113, was evaluated as a false negative. The poor coverage of the structure by the feature identified indicates that an expert review of the prediction is necessary. An expert review may consider the training set structures, which map to the feature, potentially reactive features, and analogous structures. Analogous structures with a similarity score greater than 30% are indicated. Of these, it is important to examine the analogs and identify if any (based on structural or biological similarity) would be relevant for assessing the validity of the model prediction. The analog field contains two conazoles (croconazole and fluconazole, see Table 14). Based on representation from the same class, these analogs would be considered useful for further analysis. Croconazole is indicated as positive for adverse effects to female reproductive organs and fertility, while fluconazole is negative for these effects. Accessing the underlying data for fluconazole indicates result findings of specific developmental abnormalities to the central nervous system, craniofacial, and musculoskeletal system (Lopez-Rangel & Van Allen, 2005). Such information may alert the reviewer to the lower reliability of the negative prediction and may support overturning the prediction based on review findings. The RMR model identified a kind of chlorophenol feature as a mitigating structural feature (Table 14), but no analog with at least 30% global similarity to triadimenol could be detected by the model. Therefore, the RMR model considered the pesticide to be outside the AD.

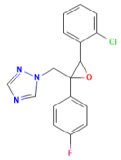
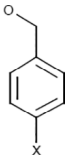
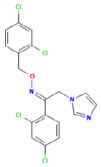
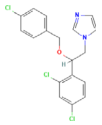
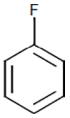
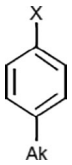
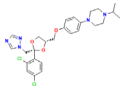
For epoxiconazole, the true positive classifications by the RFR model and false positive classification by the RMR model were based on evidence of structural feature contributions (RFR: benzene, 1-halo, 4-oxymethyl-feature, RMR: Fluorobenzene structure represents one of four identified features, see Table 15) and property descriptors associated with the predicted specific effect. Given the totality of positive/negative contributing traits in the pesticide structure, the positive probability for reprotoxicity in both models was above the cut-off: the RFR model identified a positive probability of .614 for the true-positive prediction and the RMR model for the false positive result was .514, which is slightly above the cut-off positive prediction by both Leadscope models. The structural similarity of the analogs with epoxiconazole was between 32 and 39% in both models. Looking at the identified analogs, it is striking that of the 7 (RFR) or 6 (RMR) conazole analogs, only one is positive for the respective specific toxicity (see Table 15). This could mislead to the unreflecting assumption that both predictions are wrong, although this is only true for the RMR prediction. Therefore, this information must be carefully considered in the context of an expert opinion.

Hence, the low reliability of the model predictions suggests that an expert review is necessary in predicting reprotoxic conazole fungicides within a chemical class that is mainly negative for toxic effects on reproduction.

3.4.3 | Summary

The conazole case study illustrates quite well the general problems of the QSAR prediction for pesticides using the Leadscope software. One of the main issues is that the identified structural features only cover part of the pesticide molecule. In the case of the conazoles, mainly benzene structures were identified. Therefore, an expert review is recommended, especially in the case of poor structural coverage. Relevant analog structures (inside and outside the Leadscope database) should also be taken into account. Additionally, it is important for the

TABLE 15 Detected structural features and selected training set analogs of epoxiconazole, which is reprotoxic in female rats

Predicted pesticide	Model	Evaluation	Detected structural features	Selected relevant analog structures	
Epoxiconazole CAS no. 135319-73-2 	RFR v2	TP	Benzene, 1-halo-, 4-oxymethyl- 	*Oxiconazole* *Positive* for RFR 	*Econazole* *Negative* for RFR 
	RMR v2	FP	Benzene, 1-fluoro-  Benzene, 1-alkyl-,2-halo-  Benzene, 1-alkyl-,2-chloro-  Benzene, 1-alkyl-,4-halo- 	*Terconazole* *Positive* for RMR 	*Econazole* *Negative* for RMR For structure, see above

assessment to confirm that the probability score is not directly related to the reliability of the prediction. Overall, a majority of the pesticides fell outside the AD of the model. This is due to the fact that for many pesticides not all property descriptors, not at least one structural feature and/or not at least one analogous substance could be identified in the training set.

3.5 | CASE ultra

In the following, the predictions of four selected CASE Ultra models (see Section 2.2.5) are examined. All models are statistically based structural alert models that use different data sets based on the respective endpoint. The classification is based on the alerts from which the probability is calculated. If an alert is assigned to the pesticide, the prediction can be positive or inconclusive. If there is no alert, the prediction is negative or out of domain. A known positive or known negative prediction can occur in both cases.

3.5.1 | Evaluation

The proportion of pesticides for which no prediction could be made (UNKNOWN), because they were either outside the AD of the model (out of domain) or the data were inconclusive (inclusive), was between 29 and 67% depending on the model (see Table 16 and Table S14). As a result, between 17 and 67% of reprotoxic pesticides were not recognized (see Table S14).

With the CASE Ultra models, there is also the case that tested pesticides also appear in the training data set of the respective model. This is then referred to as known positive/negative in the prediction. With the FDYSM_Rat and the MFRET_Rat model, 10 or 5 of these pesticides are nevertheless incorrectly predicted, which suggests a

TABLE 16 The results of evaluation of the four tested CASE Ultra models FDYSM_Rabbit, FDYSM_Rat, FFERT_Rat and MFERT_Rat via typical parameters

Model	# FN	# FP	# TN	# TP	# UNKNOWN	SEN	SPC	BA	ACC
FDYSM_RABBIT	4	18	75	5	208	0.56	0.81	0.68	0.78
FDYSM_RAT	19	75	115	10	91	0.34	0.61	0.48	0.57
FFERT_RAT	2	18	132	1	157	0.33	0.88	0.61	0.87
MFERT_RAT	2	62	106	3	137	0.60	0.63	0.62	0.63

Abbreviations: ACC, accuracy; BA, balanced accuracy; FN, false negative; FP, false positive; SEN, sensitivity; SPC, specificity; TN, true negative; TP, true positive.

different data basis or interpretation of the data (see Table S15). Since the CASE Ultra models cannot access the underlying data, no further investigations were carried out.

All four CASE Ultra models showed a lower sensitivity (between 0.33 and 0.6) than specificity (between 0.61 and 0.87). The FDYSM_Rat model was particularly noticeable due to its high number of false negatives (19), 6 of which belonged to the triazoles. Since there is no external validation available for the CASE Ultra models, no comparison was possible.

The reprotoxicity is determined in the CASE Ultra models using the "Probability" value. The larger the value, the more reliable a positive prediction should theoretically be. However, this is not the case in any of the models, as can be seen in Figure S4.

3.5.2 | Alerts

The determination of the reprotoxicity of the CASE Ultra models is based on statistical structural alerts. These differ between the models. If no alert fits, the prediction is limited to known positive/negative, negative and out of domain. Otherwise, all predictions are possible, including a negative one. Several alerts are possible for each pesticide, but overall, no alert was assigned to over 60% of the pesticides for all models (see Table S16). Figure S5 shows the distribution of FN, FP, TN TP, and UNKNOWN per alert and model. The problem with the alerts used is that they are often very general and only cover very small sections of the molecule. Several alerts would always be required to cover the entire molecule, which is rarely the case. From the plot just described, therefore, it was not possible to select any alerts that would provide reliable predictions.

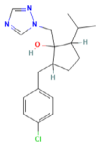
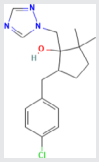
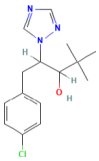
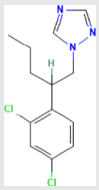
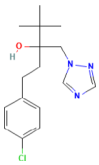
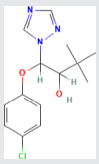
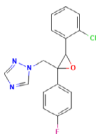
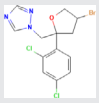
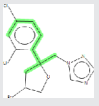
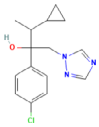
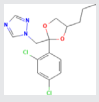
3.5.3 | Example

The prediction of reprotoxic potential of the triazoles by the FDYSM_Rat model should be used in the following to show the problems of the CASE Ultra models. The pesticide DB contains 21 triazoles of which 10 showed fetal dysmorphogenesis in rat studies relevant for ECHA classification. Three of them were predicted correctly, but six as negative and one was outside the AD of the model (see Table 17). Interestingly, the alert for all TPs was: C3H2-C3-c:cH:cH:c:cH (Alert ID 105), which describes an aromatic structure with at least one undefined substituent and a defined secondary substituent, which is a quaternary carbon followed by a secondary carbon. This alert only describes a small part of the molecule which is probably not very relevant for the toxicity mechanism as three non-reprotoxic triazoles had the same alert (difenoconazole, flutriafol, myclobutanil). No alert could be assigned for the 6 FN triazoles, which indicates that there is a data gap here. Each prediction includes the 3 closest neighbors of the test chemical in the training set. In the case of the triazoles, there are some triazoles and imidazoles among these, but a similarity above 0.7 is never reached. Thus, these cannot be regarded as analog and therefore only have a limited significance.

3.5.4 | Summary

The example shows the problem of the alerts within the CASE Ultra models. These form the basis of the prediction, but often only depict a small part of the molecular structure of the pesticides. This creates a high number of FPs. On the other hand, the critical structures are sometimes not recorded, or there is no suitable alert at all for reprotoxic pesticides, although all fragments are present in the data set. When evaluating the prediction, the alerts and their relevance should always be considered. The probability increases with an increasing number of alerts (not continuously) but is otherwise not a reliable indicator for the correctness of the prediction. When evaluating, the similarity of the 3 closest neighbors should also be considered. If this is more than 0.7, the substances can be considered analogous according to the model description. Overall, the assessment of the predictions of the CASE Ultra models also requires critical questioning by reprotoxicology experts.

TABLE 17 All triazoles of the pesticide DB that showed fetal dysmorphogenesis in ECHA classification-relevant studies in rats and their prediction by the FDYSM_Rat model from CASE Ultra

Name	CAS no.	Structure	Prediction/probability/alert
Ipconazole	125225-28-7		Negative/30.3/no alert
Metconazole	125116-23-6		Negative/30.3/no alert
Paclobutrazol	76738-62-0		Negative/30.3/no alert
Penconazole	66246-88-6		Negative/30.3/no alert
Tebuconazole	107534-96-3		Negative/30.3/no alert
Triadimenol	55219-65-3		Negative/30.3/no alert
Epoxiconazole	133855-98-8		Out of domain/30.3/no alert
Bromuconazole	116255-48-2		Positive/56/alert ID 105: C3H2-C3-c:cH:cH:c:cH 
Cyproconazole	94361-06-5		Positive/56/alert ID 105: C3H2-C3-c:cH:cH:c:cH 
Propiconazole	60207-90-1		Positive/56/alert ID 105: C3H2-C3-c:cH:cH:c:cH 

Note: When an alert was found, the relevant structure in the molecular pesticide structure is highlighted in green.

3.6 | Comparison

There are several ways to compare the predictive power of the different models. The accuracy used for this is usually the one that should not be considered on its own in the case of an unbalanced training or test data set. This can be seen, for example, on the RFR model from Leadscope, which was rated the highest with an accuracy of 0.96. However, the sensitivity was only 0.25, which means that three quarters of all reprotoxic pesticides were not detected (see Figure 5 and Table S17). The balanced accuracy, which is the mean value of sensitivity and specificity, offers a better reference point. For the evaluation of the models, above all, sensitivity and specificity are decisive. Since the safety aspect plays a decisive role in predicting reprotoxicity and a low specificity is more tolerable than a low sensitivity, the focus is more on sensitivity. A typical representation for this is the ROC diagram in which the false positive rate (FPR, 1-sensitivity) is plotted against the true positive rate (TPR, sensitivity). The closer the models are to the diagonal (black line), the more the prediction resembles a random process (see Figure 5b). Another important point to consider when assessing predictive power is how many of the pesticides were within the AD of the model and given a reliable score. In the models tested, this was between 100 and 22%.

The high number of "UNKNOWN" shows clearly that the majority of the models are not suitable for predicting pesticides, as these are outside the chemical space of the models. A sensitivity above 0.55 is only achieved with four models, whereas the CAESAR model has a specificity of only 0.16. The other three models (OQTB, CU_FDYS_Rabbit and CU_MFERT_Rat) achieve a specificity of at least 0.63. According to this statistical evaluation, all models are insufficient for predicting reprotoxicity or the partial aspects.

Also, in the overall comparison of the PG model with the DART scheme of the OECD (Q)SAR Toolbox, it becomes clear that the predictions of the models differ significantly, although both are originally based on the same decision tree. A detailed discussion of all differences can be found in Section 3.3.1. The Leadscope and CASE Ultra reprotoxicity models are based on the same

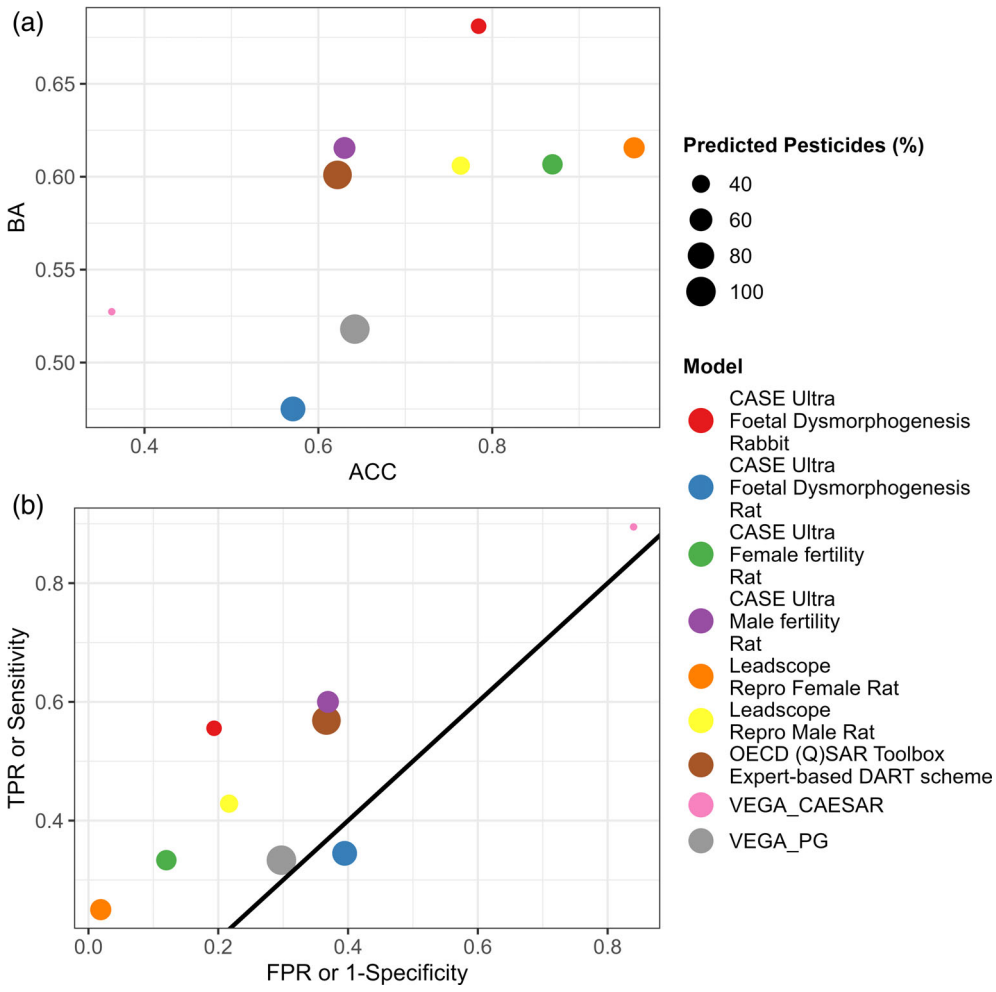


FIGURE 5 Plot of the accuracy against the balanced accuracy (a) and the FPR against the TPR (b) per model. The size of the points depends on the percentage of pesticides predicted. The black line shows the diagonal of the plot (TPR = FPR). The closer the points are to the diagonal, the more the model's prediction resembles a random process

database, but their models differ greatly (statistical QSAR vs. structural alert system), which also leads to large differences in the predictions. Both models predict toxicity of individual endpoints rather than overall reprotoxicity. However, this does not lead to an improvement in prediction reliability, as originally expected.

To find out whether the prediction quality differs between the chemical groups within the pesticide database, this was examined for the 12 largest chemical groups (see Table 2). Figure 6 shows the distribution of FP, FN, TN, TP, and UNKNOWN per chemical group for each model. These differed greatly between the models.

Of the 13 carbamates, 2 are classified as reprotoxic by the ECHA. Benfuracarb due to male reprotoxicity in the rat and carbendazim also due to male reprotoxicity in the rat but also fetal dysmorphogenesis in the rat and rabbit. When comparing the predictions of the PG model and the DART model of the OQTB for all carbamates, it is noticeable that the two reprotoxic pesticides were recognized as such, but most of the others were predicted false positives. In the PG model, almost all carbamates were also present in the training data set, which on the one hand suggests a different interpretation of the experimental data and on the other hand a general tendency of both models to classify carbamates as reprotoxic. The developmental toxicity shall be predicted by the CAESAR model. Most of the carbamates (eight pieces) were outside the model's AD and all others were predicted to be developmentally toxic, with only one actually being developmentally toxic. This phenomenon is not specific to carbamates, but in general the majority of pesticides was predicted by the CAESAR model to be developmentally

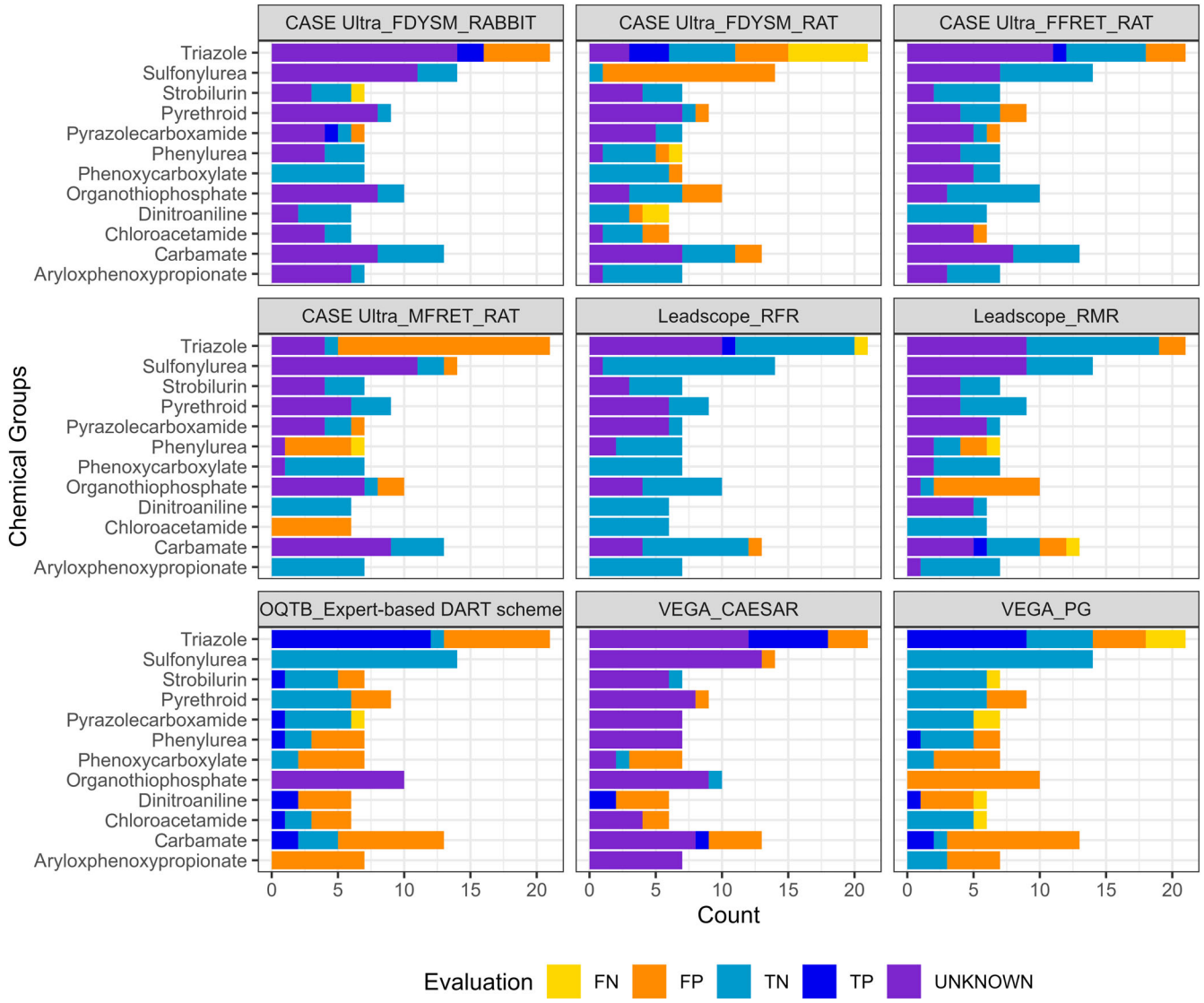


FIGURE 6 All bar plots show the distribution of FN, FP, TN, TP, and UNKNOWN per chemical group for a different prediction model. A more detailed description of the chemical groups can be found in Table 2

TABLE 18 Possible additional information on the prediction, which is made available in the reports

Information about...	Important questions
Structural alert/feature/ predicted category	Does the selected structural fragment match the key functional groups of the pesticide?
Analog structures/similar compounds from training set	How similar are these compounds?
Data sources	Which source is the classification based on? Which effects are described in this source?

toxic. In all CASE Ultra models, which each predict different aspects of reprotoxicity, more than half of the carbamates were not within AD or the prediction was inconclusive, including the reprotoxic pesticides. Only the FDYSM_Rat model predicted two carbamates as FP. In the Leadscope models, there were four and five carbamates outside the AD of the models. In the RFR model, the majority of carbamates TN and only one FP was predicted. Two reprotoxic ones were expected in the RMR model, of which carbendazim was recognized, but benfuracarb was predicted to be FN. Two carbamates were predicted in FP and four in TN.

Overall, it is noticeable that the prediction quality of the models, except for the CAESAR model, which generally tends to predict FP, differs between the chemical groups. If one compares the prediction quality of the chemical groups between the models, there are also major differences (e.g., dinitroaniline) and some models are then better suited than others for predicting the reprotoxicity of certain chemical groups.

By looking at the prediction quality in relation to selected groups, it becomes clear that the individual models can provide good predictions under certain conditions. When predicting the reprotoxicity using *in silico* models, it is therefore important to consider the predictions of several models and to weight them using the additional information provided in the report (alerts, similar compounds from the training data set) in order to arrive at a well-founded opinion. The relevant additional information that should be analyzed is summarized in Table 18.

4 | CONCLUSION

The aim of this paper was to test the performance of known models for predicting reproductive toxicity of pesticides and to use the results to analyze the strengths and weaknesses of the models. This resulted in suggested solutions for improving the models. The paper is intended to address three different target groups: *In silico* experts are to be made aware of the special problems of reprotoxicity, regulatory toxicologists are to be made aware of the limitations of the individual models and reprotoxicologists are to be made familiar with the *in silico* topic in order to point out what contribution they can still make.

The models used differed in several aspects (see Table 1 for details):

- Type of model (statistical model, expert rule-based model or mixture)
- Training data set
- Endpoint (general reprotoxicity vs. selected reprotoxicity endpoints)

However, the comparison of the models does not allow any statement to be made as to which model type, training data set or endpoint is most suitable, since all models have major weaknesses in assessing reprotoxicity of pesticides. In four of the nine models, no reliable prediction can be made for over 50% of the pesticides and in five out of nine models, not even half of the reprotoxic pesticides are recognized (SEN < 0.5). In contrast, all models except the CAESAR model recognize at least 60% of the negative pesticides. Of course, the performance of the models differs but overall, no model is convincing if all three factors (number of predicted pesticides, SEN, SPC) are taken into account.

There are three main reasons for the poor performance of the models in relation to the pesticide database:

1. Many pesticides are not part of the chemical space of the models. For example, the CAESAR model, which is based on a drug database, cannot provide a reliable prediction for more than three quarters of all pesticides. Due to its database, it is only suitable to a limited extent for predicting pesticides. In general, however, this problem is due to a too small database with high-quality reprotoxicity studies of pesticides. Therefore, larger databases based on uniform study designs would be needed to improve the models.
2. Definitions of reproductive toxicity vary. The unification of the assessment of toxicity is still a current issue for the *in vivo* area since the interpretations are also partly different here. For *in silico* toxicology, an important step here would no longer be to predict the entire reprotoxicity, but rather more easily definable, specific endpoints or effects. Even if this could not be shown with the models used, a better predictive power

can be expected from endpoint-specific models, since they are based on a smaller number of possible AOPs.

3. The partially insufficient definition of similarity within models. With the models provided by VEGA, the most similar molecules from the respective training data set are displayed in the report and used to calculate the reliability score (ADI). It is important to note that this analysis is independent of the prediction model. VEGA tended to overestimate the similarity of the structures (see Section 3.1.2 example). With the PG model and the DART scheme of the OECD (Q)SAR Toolbox, there were sometimes incorrect classifications into categories (see Sections 3.2.2 and 3.3.2). The Leadscope models include structural features, and the CASE Ultra models are alarm based. In both cases it would be desirable for the structural features or alerts to cover the entire molecular structure, but this is practically never achieved, which is more serious in the case of the CASE Ultra model. Since all prediction models, regardless of type, are essentially based on similarity, optimizing the calculation of similarity is a crucial step in improving the models. In order to describe similarity, there are more possibilities apart from fingerprints and descriptors, which should be used: AOPs, metabolism, receptor binding etc. At the structural level, the use of SMARTs or higher order substructures, that could even include metabolism information, would also be a possibility. It is crucial that the structures and properties relevant to the toxicity can be fully described using the selected parameters.

Despite all their weaknesses, the models can be of great use when used critically and the results compared to other models. Ensemble/consensus-type approaches are suitable for this, which potentially make it possible to compensate for the weaknesses of one model with another. All models provide the reasons for the prediction (alerts) and/or similar molecules from the training data set in their respective report. This information usually allows a good assessment of the plausibility of the prediction, provides clues for further research and should therefore always be analyzed carefully. The DART scheme of the OECD (Q)SAR Toolbox and the PG model occupy a special position within the tested models, since they are both based on the expert-known-based decision tree by Wu et al. (2013). This gives a good overview of chemical groups with known reprotoxicity and can serve as a starting point for the development/inclusion of MOAs and AOPs.

All the points mentioned are of course suitable for improving prediction models, regardless of the type of toxicity. For reprotox, however, the conditions are more difficult overall due to the small amount of available and high-quality data, the complexity of the underlying studies, the knowledge gaps regarding the modes of action and the point that reprotoxicity is a mixture of effects, which encompass a number of endpoints. Solving the problems just described is essential for the development of successful reprotoxicity models. Until then, using the models already available requires a critical look at the results based on reprotox expertise.

ACKNOWLEDGMENTS

The authors thank the teams of CASE Ultra, Leadscope, and the OECD (Q)SAR Toolbox for their support.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Anastasia Weyrich https://orcid.org/0000-0003-2963-5117

REFERENCES

Arena, V. C., Sussman, N. B., Mazumdar, S., Yu, S., & Macina, O. T. (2004). The utility of structure-activity relationship (SAR) models for prediction and covariate selection in developmental toxicity: Comparative analysis of logistic regression and decision tree models. *SAR and QSAR in Environmental Research*, *15*(1), 1–18. https://doi.org/10.1080/1062936032000169633

Avram, M. J. (2020). Pharmacokinetic studies in pregnancy. *Seminars in Perinatology*, *44*(3), 151227. https://doi.org/10.1016/j.semperi.2020.151227

Bantle, J. A., Fort, D. J., & James, B. L. (1989). Identification of developmental toxicants using the frog embryo Teratogenesis assay-Xenopus (FETAX). *Hydrobiologia*, *188*(1), 577–585. https://doi.org/10.1007/BF00027826

Benfenati, E. (2020). Developmental/Reproductive Toxicity Library (PG) (version 1.1.0). QMRF.

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., ... Wiswedel, B. (2008). KNIME: The Konstanz information miner. In I. C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data analysis, machine learning and applications Berlin*. Heidelberg: KNIME: The Konstanz Information Miner.

Cassano, A., & Benfenati, E. (2010). Developmental toxicity model (CAESAR) – v. 2.1.7. QMRF.

Cassano, A., Manganaro, A., Martin, T., Young, D., Piclin, N., Pintore, M., ... Benfenati, E. (2010). CAESAR models for developmental toxicity. *Chemistry Central Journal*, *4* (Suppl 1), S4. https://doi.org/10.1186/1752-153X-4-S1-S4

Cereto-Massagué, A., Ojeda, M. J., Valls, C., Mulero, M., Garcia-Vallvé, S., & Pujadas, G. (2015). Molecular fingerprint similarity search in virtual screening. *Methods*, *71*, 58–63. https://doi.org/10.1016/j.ymeth.2014.08.005

Chakravarti, S. K., Saiakhov, R. D., & Klopman, G. (2012). Optimizing predictive performance of CASE ultra expert system models

using the applicability domains of individual toxicity alerts. *Journal of Chemical Information and Modeling*, *52*(10), 2609–2618. https://doi.org/10.1021/ci300111r

Cioffi, G. (2019a). RP_FDYSM_RABBIT, fetal Dysmorphogenesis, Rabbit, Model version, 1.7.0.5.257.500. QMRF.

Cioffi, G. (2019b). RP_FDYSM_RAT, fetal Dysmorphogenesis, Rat, Model version, 1.7.0.5.892.450. QMRF.

Cioffi, G. (2019c). RP_FFERT_RAT, male fertility, Rat, Model version, 1.7.0.5.226.550. QMRF.

Cioffi, G. (2019d). RP_MFERT_RAT, male fertility, Rat, Model version, 1.7.0.5.360.500. QMRF.

Cronin, M. T. D., & Worth, A. P. (2008). (Q)SARs for predicting effects relating to reproductive toxicity. *QSAR & Combinatorial Science*, *27*(1), 91–100. https://doi.org/10.1002/qsar.200710118

Danishuddin, M., & Khan, A. U. (2016). Descriptors and their selection methods in QSAR analysis: Paradigm for drug design. *Drug Discovery Today*, *21*(8), 1291–1302. https://doi.org/10.1016/j.drudis.2016.06.013

De Schaepdrijver, L. M., Annaert, P. P. J., & Chen, C. L. (2019). Ontogeny of ADME processes during postnatal development in man and preclinical species: A comprehensive review. *Drug Metabolism and Disposition*, *47*(3), 295. https://doi.org/10.1124/dmd.118.084350

Enoch, S. J., Hasarova, Z., Cronin, M. T. D., Bridgwood, K., Rao, S., Kluxen, F. M., & Frericks, M. (2022). Sub-structure-based category formation for the prioritisation of genotoxicity hazard assessment for pesticide residues: Sulphonyl ureas. *Regulatory Toxicology and Pharmacology*, *129*, 105115. https://doi.org/10.1016/j.yrtph.2022.105115

Floris, M., Manganaro, A., Nicolotti, O., Medda, R., Mangiatordi, G. F., & Benfenati, E. (2014). A generalizable definition of chemical similarity for read-across. *Journal of Cheminformatics*, *6*(1), 39. https://doi.org/10.1186/s13321-014-0039-1

Hansch, C., & Fujita, T. (1964). P-σ-π analysis. A method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society*, *86*(8), 1616–1626. https://doi.org/10.1021/ja01062a035

Hanser, T., Barber, C., Guesné, S., Marchaland, J. F., & Werner, S. (2019). Applicability domain: Towards a more formal framework to express the applicability of a model and the confidence in individual predictions. In H. Hong (Ed.), *Advances in computational toxicology: Methodologies and applications in regulatory science* (pp. 215–232). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-16443-0_11

Hewitt, M., Ellison, C. M., Enoch, S. J., Madden, J. C., & Cronin, M. T. D. (2010). Integrating (Q)SAR models, expert systems and read-across approaches for the prediction of developmental toxicity. *Reproductive Toxicology*, *30*(1), 147–160. https://doi.org/10.1016/j.reprotox.2009.12.003

Khera, K. S. (1987). Maternal toxicity in humans and animals: Effects on fetal development and criteria for detection. *Teratogenesis, Carcinogenesis, and Mutagenesis*, *7*(3), 287–295. https://doi.org/10.1002/tcm.1770070309

King, C. T., Rogers, P. D., Cleary, J. D., & Chapman, S. W. (1998). Antifungal therapy during pregnancy. *Clinical Infectious Diseases*, *27*(5), 1151–1160. https://doi.org/10.1086/514977

Kjærstad, M. B., Taxvig, C., Nellemann, C., Vinggaard, A. M., & Andersen, H. R. (2010). Endocrine disrupting effects in vitro of conazole antifungals used as pesticides and pharmaceuticals. *Reproductive Toxicology*, *30*(4), 573–582. https://doi.org/10.1016/j.reprotox.2010.07.009

Leadscope. (2021). B-3. Reproductive Toxicity Suite. Leadscope Model Applier Documentation, *Version 3.1*.

Lopez-Rangel, E., & Van Allen, M. I. (2005). Prenatal exposure to fluconazole: An identifiable dysmorphic phenotype. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *73*(11), 919–923. https://doi.org/10.1002/bdra.20189

Matthews, E. J., Kruhlak, N. L., Cimino, M. C., Benz, R. D., & Contrera, J. F. (2006a). An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. *Regulatory Toxicology and Pharmacology*, *44*(2), 83–96. https://doi.org/10.1016/j.yrtph.2005.11.003

Matthews, E. J., Kruhlak, N. L., Cimino, M. C., Benz, R. D., & Contrera, J. F. (2006b). An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using in silico methods. *Regulatory Toxicology and Pharmacology*, *44*(2), 97–110. https://doi.org/10.1016/j.yrtph.2005.10.004

Matthews, E. J., Kruhlak, N. L., Daniel Benz, R., & Contrera, J. F. (2007). A comprehensive model for reproductive and developmental toxicity hazard identification: I. Development of a weight of evidence QSAR database. *Regulatory Toxicology and Pharmacology*, *47*(2), 115–135. https://doi.org/10.1016/j.yrtph.2006.11.002

Matthews, E. J., Kruhlak, N. L., Daniel Benz, R., Ivanov, J., Klopman, G., & Contrera, J. F. (2007). A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals. *Regulatory Toxicology and Pharmacology*, *47*(2), 136–155. https://doi.org/10.1016/j.yrtph.2006.10.001

Mellor, C. L., Marchese Robinson, R. L., Benigni, R., Ebbrell, D., Enoch, S. J., Firman, J. W., ... Cronin, M. T. D. (2019). Molecular fingerprint-derived similarity measures for toxicological read-across: Recommendations for optimal use. *Regulatory Toxicology and Pharmacology*, *101*, 121–134. https://doi.org/10.1016/j.yrtph.2018.11.002

Mogensen, D. M., Pihl, M. B., Skakkebæk, N. E., Andersen, H. R., Juul, A., Kyhl, H. B., ... Jensen, T. K. (2017). Prenatal exposure to antifungal medication may change anogenital distance in male offspring: A preliminary study. *Environmental Health*, *16*(1), 68. https://doi.org/10.1186/s12940-017-0263-z

Nitzsche, D. (2017). Effect of maternal feed restriction on prenatal development in rats and rabbits - A review of published data. *Regulatory Toxicology and Pharmacology*, *90*, 95–103. https://doi.org/10.1016/j.yrtph.2017.08.009

OECD. (2001). Test no. 416: Two-generation reproduction toxicity. OECD Publishing. https://doi.org/10.1787/9789264070868-en

OECD. (2006). Report on the regulatory uses and applications in OECD member countries of (quantitative) structure-activity relationship [(Q)SAR] models in the assessment of new and existing chemicals. OECD Papers, *vol. 6/11*. https://doi.org/10.1787/oecd_papers-v6-art37-en

OECD. (2018a). *Test no. 414: Prenatal developmental toxicity study* (Vol.). OECD Publishing. https://doi.org/10.1787/9789264070820-en

OECD. (2018b). Test no. 443: Extended one-generation reproductive toxicity study. OECD Publishing. https://doi.org/10.1787/9789264185371-en

OECD (Q)SAR Toolbox. (2020). *Description of the DART scheme, v1.4. Procter & Gamble*. Bourgas, Bulgaria: Laboratory of Mathematical Chemistry (LMC).

Piersma, A. H., Genschow, E., Verhoef, A., Spanjersberg, M. Q., Brown, N. A., Brady, M., ... Spielmann, H. (2004). Validation of the postimplantation rat whole-embryo culture test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Alternatives to Laboratory Animals*, *32*(3), 275–307. https://doi.org/10.1177/026119290403200307

R Core Team. (2019). R: A language and environment for statistical computing. In R Foundation for Statistical Computing. https://www.R-project.org/

Raies, A. B., & Bajic, V. B. (2016). In silico toxicology: Computational methods for the prediction of chemical toxicity. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *6*(2), 147–172. https://doi.org/10.1002/wcms.1240

Russell, W. M. S., & Burch, R. L. (1959). The principles of humane experimental technique. Methuen.

Seiler, A. E. M., & Spielmann, H. (2011). The validated embryonic stem cell test to predict embryotoxicity in vitro. *Nature Protocols*, *6*(7), 961–978. https://doi.org/10.1038/nprot.2011.348

Selderslaghs, I. W., Van Rompay, A. R., De Coen, W., & Witters, H. E. (2009). Development of a screening assay to identify teratogenic and embryotoxic chemicals using the zebrafish embryo. *Reproductive Toxicology*, *28*(3), 308–320. https://doi.org/10.1016/j.reprotox.2009.05.004

Škuta, C., Cortés-Ciriano, I., Dehaen, W., Kříž, P., van Westen, G. J. P., Tetko, I. V., ... Svozil, D. (2020). QSAR-derived affinity fingerprints (part 1): Fingerprint construction and modeling performance for similarity searching, bioactivity classification and scaffold hopping. *Journal of Cheminformatics*, *12*(1), 39. https://doi.org/10.1186/s13321-020-00443-6

Society for the Advancement of Adverse Outcome Pathways. (2022). Adverse outcome pathways for Reprotox. https://aopwiki.org/

Tasnif, Y., Morado, J., & Hebert, M. F. (2016). Pregnancy-related pharmacokinetic changes. *Clinical Pharmacology and Therapeutics*, *100*(1), 53–62. https://doi.org/10.1002/cpt.382

Tetro, N., Moushaev, S., Rubinchik-Stern, M., & Eyal, S. (2018). The placental barrier: The gate and the fate in drug distribution. *Pharmaceutical Research*, *35*(4), 71. https://doi.org/10.1007/s11095-017-2286-0

Theunissen, P. T., Beken, S., Beyer, B. K., Breslin, W. J., Cappon, G. D., Chen, C. L., ... Piersma, A. H. (2016). Comparison of rat and rabbit embryo-fetal developmental toxicity data for 379 pharmaceuticals: On the nature and severity of developmental effects. *Critical Reviews in Toxicology*, *46*(10), 900–910. https://doi.org/10.1080/10408444.2016.1224807

U.S. EPA. (2005a). 2,4-D HED's revised human health risk assessment for the reregistration eligibility decision (RED) revised to reflect public comments.

U.S. EPA. (2005b). 2,4-D Reregistration Eligibility Decision. (EPA 738-R-05-002). Washington, DC: U.S. EPA.

Valerio, L. G. (2009). In silico toxicology for the pharmaceutical sciences. *Toxicology and Applied Pharmacology*, *241*(3), 356–370. https://doi.org/10.1016/j.taap.2009.08.022

Venkatapathy, R., & Wang, N. C. Y. (2013). Developmental toxicity prediction. In B. Reisfeld & A. N. Mayeno (Eds.), *Computational toxicology* (Vol. II, pp. 305–340). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-62703-059-5_14

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. In Springer-Verlag.

Wu, S., Fisher, J., Naciff, J., Laufersweiler, M., Lester, C., Daston, G., & Blackburn, K. (2013). Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. *Chemical Research in Toxicology*, *26*(12), 1840–1861. https://doi.org/10.1021/tx400226u

Yang, H., Lou, C., Li, W., Liu, G., & Tang, Y. (2020). Computational approaches to identify structural alerts and their applications in environmental toxicology and drug discovery. *Chemical Research in Toxicology*, *33*(6), 1312–1322. https://doi.org/10.1021/acs.chemrestox.0c00006

Zarn, J. A., Brüschweiler, B. J., & Schlatter, J. R. (2003). Azole fungicides affect mammalian steroidogenesis by inhibiting sterol 14 alpha-demethylase and aromatase. *Environmental Health Perspectives*, *111*(3), 255–261. https://doi.org/10.1289/ehp.5785

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Weyrich, A., Joel, M., Lewin, G., Hofmann, T., & Frericks, M. (2022). Review of the state of science and evaluation of currently available *in silico* prediction models for reproductive and developmental toxicity: A case study on pesticides. *Birth Defects Research*, 1–31. https://doi.org/10.1002/bdr2.2062

4.3 How to improve prediction of teratogenicity? – A case study on anticoagulant rodenticides

4.3.1 Introduction

Although *in silico* methods are primarily used for screening, there are also substances whose classification is based on an *in silico* approach. An example of this is the reclassification of currently used anticoagulant rodenticides by the EU Commission (Commission Regulation (EU) 2016/1179) on the recommendation of ECHA. The rodenticides with an anticoagulant dose of 0.003% or higher must be labelled with the hazard symbol "toxic for reproduction" and are also prohibited for amateur use under the Biocidal Products Regulation (BPR, Regulation (EU) 528/2012). This includes: chlorophacinone, warfarin, coumatetralyl, difenacoum, brodifacoum, flocoumafen, bromadiolone and difethialone. Except for warfarin, none of the rodenticides were previously classified as reprotoxic because the underlying developmental toxicity studies in rats and rabbits showed no developmental toxicity. The new classification was based on a read-across approach, which assumes that the reprotoxicity is the same due to an equal MoA.

Read-across approaches are about filling data gaps using similar molecules with known toxicity. This can be partly automated, for example with the help of the OECD QSAR Toolbox, or based on expert knowledge. In any case, however, there is a great amount of time involved, mainly due to compilation and preparation of the database. With QSAR models, on the other hand, the prediction is automated after the model has been created and can ideally be applied to many substances. As explained in the previous paper, a weakness of current QSAR models for *in silico* toxicology is that they only use a structure within the molecule to predict toxicity. In contrast to the read-across approaches, information such as the MoA, AOP or AMDE data is currently not used in QSAR models for developmental toxicity. The following case study is intended to use the example of anticoagulant rodenticides to show which additional information can be used in QSAR models and what benefit results from this.

4.3.2 Coumarins - Definition and Use

4-Hydroxycoumarins are vitamin K antagonists (VKAs) and belong to the anticoagulants. Because of this feature, they find application as drugs and rodenticides. Their effect is based on the competitive inhibition of vitamin K epoxide reductase (VKOR) and leads to the shutting down of vitamin K cycle, which is essential for the function of vitamin K-dependent proteins (VKDPs) like blood coagulation factors II, VII, IX and X.

4.3.2.1 Vitamin K and vitamin K-dependent proteins

Vitamin K is a fat-soluble and essential vitamin, which acts as cofactor for γ-glutamyl carboxylase (GGCX). These activates the VKDPs by carboxylating the glutamic acid (Glu) residues of VKDPs to gamma-carboxyglutamate (Gla). Due to the high need of vitamin K, it is recycled via the vitamin K cycle (see figure 5)[117]: In the target cells vitamin K is converted

from a stable oxidized form (quinone form) to a hydroquinone form by vitamin K epoxide reductase (VKOR). This acts as a cofactor of GGCX, which carboxylates the VKDPs, while simultaneously oxidizing the reduced form of vitamin K to an epoxide form. Conclusive the epoxide form is reduced by epoxide reductase to the original vitamin K.

The VKDPs also called Gla-proteins can be classified due to their main actions [118]:

1. Involved in blood coagulation
   Procoagulant: Factor II (Prothrombin), VII (Proconvertin), IX, and X
   Anticoagulant: Protein C, S, and Z
2. Connective tissue mineralization
   Matrix Gla protein (MGP), osteocalcin (BGP, bone Gla protein), Gla-rich protein, and nephrocalcin
3. Transmembrane receptors
   Transmembrane Gla proteins 3 and 4 (TGM3, TGM4), Proline-rich Gla proteins (PRGP1, PRGP2)
4. Other effects
   Gas6 (growth factor) and others

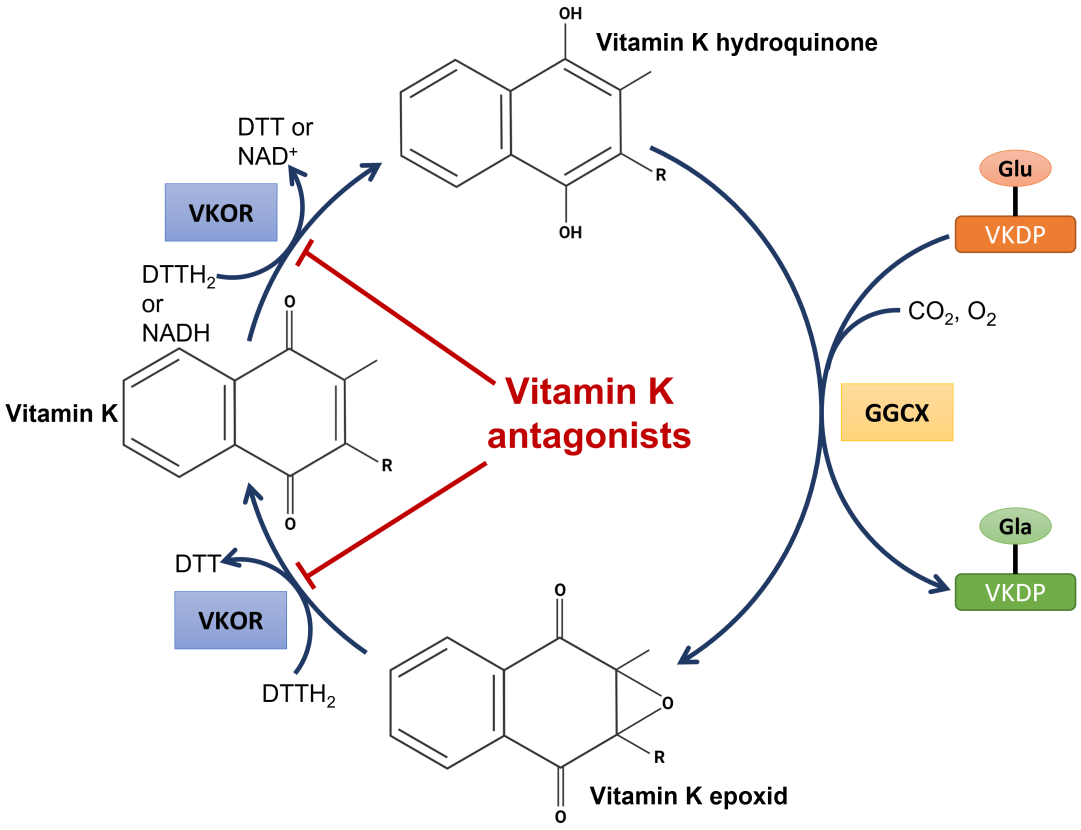


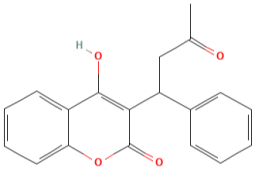
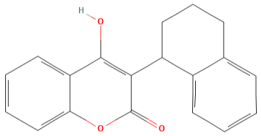
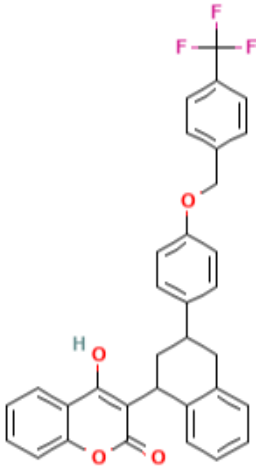
Figure 5: Overview of vitamin K cycle with attack points of vitamin K antagonists. Figure was designed according to Hirota & Suhara (2019) [119]. VKOR: Vitamin K epoxide reductase; GGCX: γ-glutamyl carboxylase; VKDP: vitamin K-dependent protein; DTT: Dithiothreitol.

4.3.2.2 Groups

The first 4-hydroxycoumarin discovered and characterized was dicoumarol ([120]). Based on the structure of 4-hydroxycoumarin, both anticoagulant drugs and rodenticides were developed from then on. Warfarin, which have a higher anticoagulant potency, was the first registered rodenticide. Development of resistance to warfarin in rat populations led to the development of the 2nd generation 4-hydroxycoumarin-based rodenticides, like brodifacoum, bromadiolone, chlorophacinone, difenacoum, diphacinone and flocoumafen (see table 7).

The 2nd generation rodenticides are based on the same basic structure as warfarin but differ in the side chain (see table 7). These are more lipophilic, and the rodenticides are therefore more potent and have a longer duration of action. This has the advantage that, in contrast to warfarin, a single feeding is usually sufficient to achieve the deadly effect. The higher potency and efficacy is based on several factors: greater affinity for VKOR, hepatic accumulation, and unusually long biological half-lives due to high lipid solubility and enterohepatic recycling [121]. The second generation also includes anticoagulants based on 1,3-indanedione, such as diphacinone and chlorphacinone.

Table 7: Variety of VKAs used as rodenticides or drugs. The illustrations of the molecular structure are by PubChem [122].

Name	CAS No.	Structure	Usage	Classification					
				Source	Hazard Class and Category Code	Hazard Statement Code	Specific Concentration limit	Animal developmental toxicity studies	Human developmental toxicity
Warfarin	81-81-2		Drug/ Rodenticide (1st Gen)	GHS, EU	Repr.1A	H360D	C ≥ 0,003 %	Positive	YES
Coumatetralyl	5836-29-3		Rodenticide (1st Gen)	GHS, EU	Repr.1B	H360D	C ≥ 0,003 %	Negative	No data
Flocoumafen	90035-08-8		Rodenticide (2nd Gen)	CLP notifications	-	-	-	Negative	No data

Brodifacoum	56073-10-0		Rodenticide (2nd Gen)	GHS, EU	Repr.1A	H360D	C ≥ 0,003 %	Negative	YES (2 human cases which indicates FWS)
Bromadiolone	28772-56-7		Rodenticide (2nd Gen)	GHS, EU	Repr.1B	H360D	C ≥ 0,003 %	Negative	No data
Difenacoum	56073-07-5		Rodenticide (2nd Gen)	GHS, EU	Repr.1B	H360D	C ≥ 0,003 %	Negative	No data

Difethialone	104653-34-1		Rodenticide (2nd Gen)	GHS, EU	Repr.1B	H360D	C ≥ 0,003 %	Negative	No data
Diphacinone	82-66-6		Rodenticide (2nd Gen)	GHS, EU	-	-	-	Negative	No data
Chlorophacinone	3691-35-8		Rodenticide (2nd Gen)	GHS, EU	Repr.1B	H360D	C ≥ 0,003 %	Negative	No data

4.3.2.3 Teratogenicity

The foetal warfarin syndrome (FWS) describes the developmental toxic properties of warfarin, first described by Hall et al. [123]. The classic features of FWS are teratogenic findings like nasal hypoplasia and epiphyseal and vertebral stippling [124]. In addition, skeletal and other malformations as well as optic atrophy and nervous system (CNS) abnormalities occur frequently. This is usually accompanied by foetal anaemia. The severity and extent are variable and depend on the dose and period of exposure during pregnancy.

Evidence of the developmental toxicity of warfarin comes from clinical reports of pregnant women treated with warfarin as an anticoagulant drug. In these reports, the typical effects mentioned were described. The teratogenicity of warfarin could also be shown in animal studies according to the OECD 414 guideline in rats. The occurring effects (intrauterine death of foetuses, increased post implantation loss, internal and subcutaneous haemorrhages, intracerebral haematomas, cataract of lens) are similar to those in men with the exception of nasal hypoplasia, which does not occur in rats. In addition to warfarin, other VKAs like phenprocoumon used as anticoagulant drugs also show the typical symptoms of FWS in men [123, 125].

The developmental toxicity of 2nd generation rodenticides is controversial because, unlike those with warfarin, associated animal studies have been negative. Nevertheless, due to the joint MoA, ECHA classified them as toxic for reproduction. This was also done on the grounds that in the animal studies with 2nd generation rodenticides too low doses were used to be able to show the teratogenic effect due to the high maternal toxicity.

Little is known about the adverse outcome pathways (AOPs) on which the different effects are based. There is only one AOP in the AOP Wiki that refers to VKAs: "187: Anticoagulant rodenticide inhibition of vitamin K epoxide reductase resulting coagulopathy and hemorrhage". However, this AOP document is still under development. It is scientifically accepted that the reproductive toxic effects are due to the same MoA and thus the same molecular initiation event (MIE): the competitive inhibition of the VKOR. The probable AOP for forming the FWS can be seen in figure 6. The inhibition of the vitamin K cycle means that the gamma-glutamyl carboxylation of the VKDPs (KE1) cannot take place, which is essential for their function. The KEs that now follow have hardly been explored to date. However, individual effects can be associated with specific VKDPs:

- Inactivity of blood coagulation factors (Factor II (Prothrombin), VII (Proconvertin), IX, and X) leads to foetal anaemia [123].
- Matrix Gla protein and osteocalcin are known for their activity in the regulation of mineralisation and bone formation, whereas the mechanism was not fully understood.

It is therefore suspected that the skeletal abnormalities typical of FWS are due to the inactivity of these two proteins [123, 124, 126, 127].

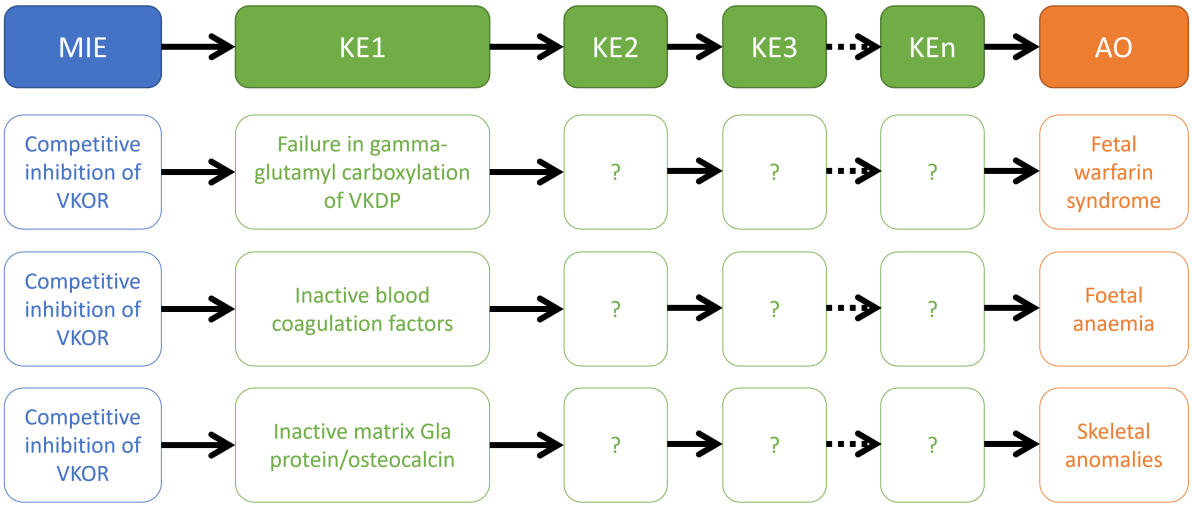


Figure 6: Presumed AOP for the forming of FWS.

4.3.2.4 ADME

In addition to the MoA, the toxicokinetics play a decisive role in the development of the toxicity. As, this is the main difference in the anticoagulant effect of warfarin to the 2nd generation rodenticides [128], it can also be the reason for the differences in the developmental toxicity in animal studies.

After mostly oral exposure, absorption is generally good via the gastrointestinal tract. This is followed by accumulation and binding in the liver, which is more pronounced in 2nd generation rodenticides than in 1st generation rodenticides, presumably due to its high lipophilicity [129]. There are also differences between the rodenticides when it comes to metabolization in the liver. Warfarin is metabolized in the liver to 6-, 8-, and 7-hydroxywarfarin, while flocoumafen for example is excreted largely unmetabolized [130].

Accumulation in the liver results in high liver concentrations for a prolonged period. In contrast, plasma concentrations tend to decrease much more rapidly for all rodenticides. The course of the plasma concentration differs greatly between the individual rodenticides [131]. The half-life of first-generation rodenticides such as warfarin is usually significantly shorter than that of second-generation rodenticides. The long elimination half-lives are probably due, among other things, to the enterohepatic circulation that occurs with second-generation rodenticides [121].

When considering developmental toxicity, the exposure of the foetus must be considered. Warfarin is already known to be transported across the placenta [132]. It is assumed that the other VKA rodenticides are also able to cross the placental barrier, but detailed studies for all

rodenticides do not yet exist. The distribution of warfarin and bromadiolone in rats over time and between the dam and foetuses at GD 20 was reported by Chetot et al. 2020, examined to elucidate the causes of the different results of the teratogenicity studies [133, 134]. The paper showed that the transmission from the mother to the foetus differs greatly. Warfarin was distributed approximately half between the liver of the mother and the foetus, while bromadiolone was found almost exclusively in the liver of the mother and in very small amounts in the foetus. This has been attributed to the level of rodenticide circulating in the blood, which is the only portion available to the foetus via the placenta and is very low for bromadiolone compared to warfarin. From this, Chetot et al. concluded that this difference in pharmacokinetics is responsible for the difference in the observed teratogenic effect.

4.3.3 Solution approaches and discussion

The literature search has shown that there are many differences in toxicokinetics between warfarin and the 2nd generation rodenticides besides the common MoA. These are hardly taken into account in the conservative read-across approach of the RAC, as there are still many uncertainties in this area and validated information is not available for all rodenticides. However, as there is currently no substitute for the anticoagulant rodenticides with comparable efficacy and less toxicity, the determination of teratogenicity is needed.

Figure 7 shows how, based on the data obtained so far, an *in silico* model could be built that considers more information than the current read-across approach. In general, all known information should be used for such a model

1. Human data are the gold standard but are rarely available. For warfarin, which is used as a drug, there are case studies describing known cases, but no systematic studies on teratogenicity. In contrast, there are hardly any human case studies for VKAs, which are used exclusively as rodenticides. If there are, however, the time and amount of exposure are unknown and difficult to determine. In general, the validity of human data based on case studies must be critically questioned since for example background diseases are not known.
2. The evaluation of the teratogenicity of biocides is usually based on animal experiments, as these have been established over decades as reliable predictive models. In the case of VKAs, the problem lies in the high maternal toxicity of second-generation rodenticides, which only allows testing of low doses in the developmental toxicity study. As a result, developmental toxic effects might not be detected. The results of the acute toxicity study are suitable for estimating the magnitude of the anticoagulant effect. In general, when using data from animal studies for *in silico* prediction models, care should be taken to ensure that nomenclature is consistent, and databases are endpoint specific.

3. The support of *in silico* predictions by AOPs is a new field that offers many possibilities [135]. For this the AOP ideally would be fully established, which is not done for anticoagulant rodenticides. As shown in figure 6, the competitive inhibition of VKOR is the MIE, followed as KE1 by the failure of gamma-glutamyl carboxylation of VKDPs. Depending on which VKDP is considered, other KEs follow, which ultimately also lead to different AOs. However, little is known about the exact causes of the FWS-typical effects (AOs), apart from anaemia. For the *in silico* prediction model, the strength of inhibition of VKOR can be used on the one hand. Comparative data between 8 rodenticides using liver microsomes from rats are already available [136]. Similar experiments would also be useful for liver microsomes from men and rabbit to make a species comparison. On the other hand, the KE1, i.e. the loss of activity of the GGCX can be used. A number of *in vitro* screening assays are already available to determine this activity [137].
4. Toxicokinetics has been less considered in the regulatory assessment of rodenticides. As already described in the first publication, it can have a major influence on toxicity and can also lead to species differences. In relation to anticoagulant rodenticides, important aspects are plasma concentration versus liver accumulation over time, as well as metabolism and transfer across the placenta and thus foetal exposure. In the previous section, the results of the searched *in vivo* and *ex vivo* studies on these aspects were summarised. However, systematic studies comparing the data of all relevant rodenticides are still lacking. To fill this data gap, a variety of *in vitro* and *in silico* models are available on placental transfer, pesticide metabolites and PBPK modelling [81, 138-141]. The consideration of toxicokinetics in the foetus both *in vivo* and *in vitro* and *in silico* remains problematic due to the lack of background knowledge.

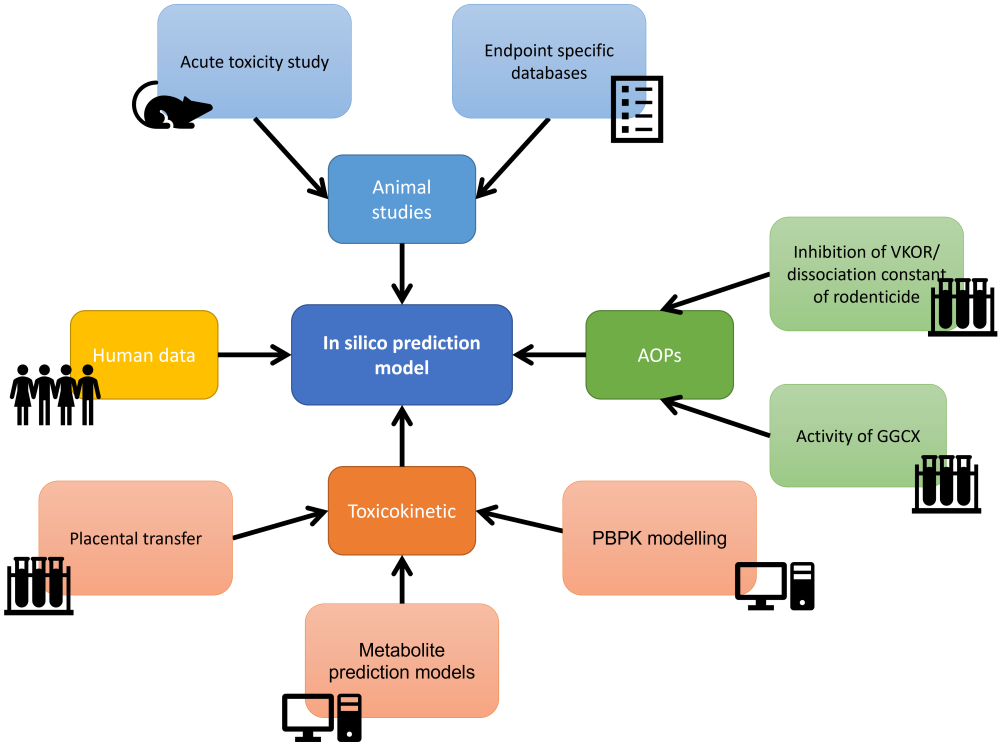


Figure 7: Summary of all aspects that should be included into an in silico prediction approach for teratogenic effects of anticoagulant rodenticides

In the case study, the example of anticoagulant rodenticides was intended to show which additional information can be used in QSAR models. Currently, the information used in *in silico* models for reprotoxicity is limited to molecular structures associated with a toxic effect. The principle is that similar structures lead to similar effects. The example of anticoagulant rodenticides shows a weakness of this system: the second-generation anticoagulant rodenticides are based on two different backbones, either 4-hydroxycoumarin or 1,3-indanedione. So, structurally the current models cannot recognise any similarity, which is why, for example, information on the dissociation constant with respect to the VKOR would be valuable information for correctly predicting toxicity. The example shows that for toxicity prediction a substance cannot be fully described by its structure alone.

The additional parameters just described, such as inhibition of VKOR, activity of GGCX, placental transfer, metabolites, and determination of kinetics are all based on *in vitro* or *in silico* assays. Compared to *in vivo* methods, these have the advantage that a much greater throughput can be achieved and thus the number of substances that flow into the model is greatly increased. Overall, substance properties that are used in addition to the structure in the development of the models can greatly improve the reliability of the models. However, there is still a great need for research on these parameters, especially in the area of reprotoxicity, since validated PBPK models that also include the placenta and the foetus, as well as AOPs that can be used to identify targets for screening assays, are lacking.

5 Final discussion and conclusion

Within toxicology, reproductive toxicology is a highly relevant field and represents a socially particularly sensitive area. It covers all toxicological events within the reproductive cycle, beginning with parental gametogenesis, through the prenatal phase from fertilisation to birth and then the postnatal phase until full sexual maturity is reached [2]. Therefore, reprotoxicity can include a variety of findings, from infertility to malformations such as spina bifida. Moreover, reproductive toxicology is very complex, as changes are constantly taking place in the mother, the placenta, and the embryo/foetus. Exposure to potentially toxic substances can occur prenatally via the placenta and postnatally via breast milk or contaminated food. This is particularly problematic as the developing organism is often more sensitive to the toxic effects of chemicals than adults due to limited detoxification mechanisms [142].

The assessment of reprotoxicity of pesticides, biocides and chemicals is based on the results of animal experiments with rats and rabbits. These are carried out in Europe according to the OECD guidelines [35, 37, 39, 41, 43, 45]. The performance and evaluation of these studies is demanding and therefore requires a very experienced laboratory team as well as study leaders. This is due to the large number of animals and long study duration and the associated quantity of endpoints to be documented. Interpretation of findings in term of classification as variations and malformations as well as the evaluation of the influence of maternal toxicity is particularly challenging. All this can lead to different reprotoxic assessment and classification of a substance despite uniform study designs.

The principle of assessing human toxicity based on *in vivo* studies in experimental animals is based on the assumption that the chosen species are sufficiently similar to men. However, as the case of thalidomide shows, there are species differences between rats, rabbits and men that are relevant for reprotoxicity [31]. The first aim of this dissertation was to investigate these species differences, focusing on in the expression of xenobiotic transporters during ontogeny.

Xenobiotic transporters are known to transport exogenous substances in addition to their endogenous substrates, as they have a very broad substrate specificity. The first xenobiotic transporter discovered, which was described by Juliano and Ling in 1976, was Mdr1, also known as P-glycoprotein [143]. It plays an important role in mediating cancer cell resistance to various cytotoxic anticancer drugs and was later assigned to the ATP-binding cassette transporter (ABC) superfamily. Another transporter superfamily known for its role in xenobiotic distribution is the solute carrier (SLC) family. The properties and functions of xenobiotic transporters are well characterised in men, rats and mice for a variety of transporters [144, 145]. In the rabbit, however, little is known about the expression, functionality, and activity of xenobiotic transporters. This is due to the fact that the rabbit is rarely used at universities and is therefore little researched in molecular biology.

Xenobiotic transporters are expressed in liver, kidney, and placental cells, among others, and can play an important role in toxicokinetics. Species differences in the kinetics of substances can have a great influence on the toxic effect. Therefore, this was the focus of the first publication. In this publication, the expression of xenobiotic transporters during ontogenesis at the mRNA level in the liver, kidney and placenta of rats and rabbits was investigated and compared with that of men. For most transporters, expression increased over ontogeny to adult levels, which is associated with increasing organ functionality. The publication showed that there are large differences in the expression of the transporters between the species, but also transporters whose expression is very similar in two or three of the species compared. It was not found that one of the two experimental animal species shows greater differences or similarities to men than the other. So neither, the rat nor the rabbit is a better translation model for xenobiotic transport in the liver, kidney or placenta compared to men. Species differences in the expression of the xenobiotic transporters may lead to differences in the distribution of their respective substrates and thus also have an impact on the potential toxicity of these substrates. However, due to the broad substrate specificity of the xenobiotic transporters and overlapping substrate spectra between the isoforms, it may be that the low transport activity of one transporter is compensated by another isoform. In order to fully assess the kinetic impact of the observed species differences, further studies on the functionality and activity of the xenobiotic transporters are required, which are explained in section 2.3 Outlook. Overall, the publication provides a valid starting point for further systematic studies of species differences at the protein level. In addition, it provides previously unavailable data on the expression of xenobiotic transporters during ontogeny in rabbits, which represents an important step in the molecular biological study of these species.

The second part focused on *in silico* prediction models for reproductive toxicology. Their predictive power was tested in relation to a pesticide database of 310 pesticides, using the parameters of sensitivity, specificity and number of pesticides predicted. Both the commercial and the freely available models did not perform adequately in the assessment. In order to determine the causes, the data basis of the models was considered on one hand and the prediction reports of false positive or false negative predicted pesticides were analysed on the other.

Overall, for a large number of pesticides (up to 77 %) no prediction was made by the models, as the pesticides were outside the chemical space of the models. This is particularly striking for models that are based solely on drug data and shows that the data basis of the models is not suitable for the prediction of pesticides. Larger databases including pesticides would be needed to improve the models. Although data from animal studies are available for all pesticides that are or were approved, they are not available in a machine-readable format. In addition, the guidelines according to which the studies were conducted have changed over

time, which may influence the results and is mostly not taken into account in the databases. Another problem is inconsistent nomenclature and terminology. By definition, approved pesticides are not allowed to be reprotoxic, and a model based on these data would contain mainly non-reprotoxic molecules, which would lead to an underestimation of reprotoxicity in the prediction. To improve the predictive power of the models for pesticides, it would be necessary to transfer the already available data into a machine-readable format with a uniform nomenclature and also to include data from developmental substances that may never have been approved precisely because of a reprotoxic effect, in order to increase the pool of positive substances.

Another problem was the different definition of reprotoxicity between the model databases and the GHS reference, reflecting the disagreement in the interpretation of the *in vivo* studies but possibly also guideline changes. An important step would be to no longer predict overall reprotoxicity as one endpoint, but rather as more definable, specific endpoints or findings, as reprotoxicity combines a large number of endpoints that are also based on different mechanisms of toxicity.

The third problem that emerged was the partly insufficient definition of similarity within the models. This has different causes and approaches depending on the model type. The QSAR-based models use a mixture of 1D descriptors and fingerprints to determine the similarity between molecules. The current algorithms tend to overestimate similarity and should be further optimized. SA-based models showed the problem that the alerts often only cover fractions of the tested molecular structure. Thus, possibly toxically relevant regions may not be considered for the prediction. This could be improved by larger and chemically diverse databases as a basis for the statistical models, as this would result in a larger set of SAs. Furthermore, SAs are context-dependent. It must therefore always be checked whether an SA is relevant in the molecular context of the test substance, e.g., by identifying specific analogues.

Ultimately, the question arises whether models that attempt to predict toxicity only based on structural similarity can be successful at all. Therefore, the use of alternative descriptors, such as biological activity or metabolism, is another approach to improve the models. These have the advantage that they can be obtained from *in vitro* tests or independent *in silico* models. However, the basis for such descriptors is knowledge about the MoA of the substance, and in the case of reproductive toxicology there is a large knowledge gap. Despite all weaknesses, the models can be of great use if their predictions are critically evaluated based on the reports. Here, all available information such as the reasons for the prediction (SAs) and/or similar molecules from the training data set should be considered. In addition, it can be useful to

compare several models using an ensemble/consensus type approach and thus increase the predictive power.

The aim of this dissertation was to develop approaches for a better assessment of reprotoxicity, with two different priorities: 1. species differences during ontogeny in relation to xenobiotic transporters and 2. reliability of the prediction of reprotoxicity by *in silico* models based on pesticides. In the first part, it was shown that species differences exist at the mRNA level. Overall, however, too little is known about the mechanisms that lead to reprotoxicity effects to be able to make reliable statements about the relevance of species differences. The evaluation of *in silico* models for reprotoxicity revealed that they are not suitable for use with pesticides. The most important approaches to solve this problem are the expansion of the databases on reprotoxicity in relation to pesticides, the improvement of the similarity assessment, the use of endpoint-specific models and the use of alternative descriptors. Overall, the dissertation shows how important it is to further investigate the modes of action of reprotoxicity in order to improve the validity of both *in vivo* studies and *in silico* models.

References

1. Pamies, D., et al., *Chapter 11 - Mechanism-based models in reproductive and developmental toxicology*, in *Reproductive and Developmental Toxicology*, R.C. Gupta, Editor. 2011, Academic Press: San Diego. p. 135-146.
2. European Medicines Agency, *ICH S5 (R3) guideline on reproductive toxicology: Detection of Toxicity to Reproduction for Human Pharmaceuticals*. 2020.
3. Hofmann, T., *Reproductive Toxicology*, in *Drug Discovery and Evaluation: Safety and Pharmacokinetic Assays*, H.G. Vogel, et al., Editors. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 1317-1327.
4. Schöni-Affolte, F., C. Dubuis-Griede, and E. Strauch. *Embryogenese*. [cited 21.06.2022]; Available from: https://embryology.ch/de/embryogenese/.
5. Ulfig, N. and B. Brand-Saberi, *Kurzlehrbuch Embryologie*. 2017, Stuttgart: Thieme.
6. Hill, M.A. *Timeline human development*. Embryology [cited 21.06.2022]; Available from: https://embryology.med.unsw.edu.au/embryology/index.php/Timeline_human_development.
7. Hill, M.A. *Rat Development*. Embryology [cited 21.06.2022]; Available from: https://embryology.med.unsw.edu.au/embryology/index.php/Rat_Development.
8. Hill, M.A. *Carnegie Stage Comparison*. Embryology [cited 21.06.2022]; Available from: https://embryology.med.unsw.edu.au/embryology/index.php/Carnegie_Stage_Comparison.
9. de Kretser, D.M., et al., *Spermatogenesis.* Human Reproduction, 1998. **13**(suppl_1): p. 1-8 DOI: 10.1093/humrep/13.suppl_1.1.
10. Gilbert SF, *Oogenesis*, in *Developmental Biology*. 2000, Sinauer Associates: Sunderland (MA).
11. Codaccioni, M., F. Bois, and C. Brochot, *Placental transfer of xenobiotics in pregnancy physiologically-based pharmacokinetic models: Structure and data.* Computational Toxicology, 2019. **12**: p. 100111 DOI: https://doi.org/10.1016/j.comtox.2019.100111.
12. Haycock, G., *CHAPTER 2 - NEPHROUROLOGY FROM FETUSHOOD TO ADULTHOOD*, in *Pediatric Urology (Second Edition)*, J.P. Gearhart, R.C. Rink, and P.D.E. Mouriquand, Editors. 2010, W.B. Saunders: Philadelphia. p. 11-22.
13. Ring, J.A., et al., *Fetal hepatic drug elimination.* Pharmacology & Therapeutics, 1999. **84**(3): p. 429-445 DOI: https://doi.org/10.1016/S0163-7258(99)00046-7.
14. Suchy, F.J., *Functional Development of the Liver*, in *Liver Disease in Children*, F.J. Suchy, R.J. Sokol, and W.F. Balistreri, Editors. 2007, Cambridge University Press: Cambridge. p. 14-27.
15. Aumüller, G., et al., *Duale Reihe Anatomie*. 2020: Thieme.
16. Carter, A.M., *Animal Models of Human Placentation – A Review.* Placenta, 2007. **28**: p. S41-S47 DOI: https://doi.org/10.1016/j.placenta.2006.11.002.
17. Carter, A.M. and A.C. Enders, *Placentation in mammals: Definitive placenta, yolk sac, and paraplacenta.* Theriogenology, 2016. **86**(1): p. 278-287 DOI: https://doi.org/10.1016/j.theriogenology.2016.04.041.
18. Carter, A.M. and A.M. Mess, *Mammalian Placentation: Implications for Animal Models*, in *Pathobiology of Human Disease*, L.M. McManus and R.N. Mitchell, Editors. 2014, Academic Press: San Diego. p. 2423-2442.
19. Freyer, C. and M.B. Renfree, *The mammalian yolk sac placenta.* J Exp Zool B Mol Dev Evol, 2009. **312**(6): p. 545-54 DOI: 10.1002/jez.b.21239.
20. Furukawa, S., Y. Kuroda, and A. Sugiyama, *A comparison of the histological structure of the placenta in experimental animals.* J Toxicol Pathol, 2014. **27**(1): p. 11-8 DOI: 10.1293/tox.2013-0060.
21. Hood, R.D. and R.D. Hood, *Handbook of Developmental Toxicology*. 1996: Taylor & Francis.
22. Giacoia, G.P. and D.R. Mattison, *Obstetric and Fetal Pharmacology.* Glob. libr. women's med., 2009 DOI: DOI 10.3843/GLOWM.10196.
23. Kalra, A., et al., *Physiology, Liver*, in *StatPearls*. 2022, StatPearls Publishing Copyright © 2022, StatPearls Publishing LLC.: Treasure Island (FL).
24. Parkinson, A., et al., *Biotransformation of Xenobiotics*, in *Casarett and Doull's Toxicology: The Basic Science of Poisons, Eighth Edition*. 2012, McGraw-Hill Education: New York, NY.
25. Godoy, P., et al., *Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mechanisms of hepatotoxicity, cell signaling and ADME.* Arch Toxicol, 2013. **87**(8): p. 1315-530 DOI: 10.1007/s00204-013-1078-5.
26. Ramadori, G. and K.-H. Meyer zum Büschenfelde, *Physiologie und Pathophysiologie der Leber und Gallenwege*, in *Pathologie der Leber und Gallenwege*. 2000, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 52-98.

27. Wells, P.G. and L.M. Winn, *5.05 - The Role of Biotransformation in Developmental Toxicity☆*, in *Comprehensive Toxicology (Third Edition)*, C.A. McQueen, Editor. 2018, Elsevier: Oxford. p. 63-85.
28. Pelkonen, O., *Biotransformation of xenobiotics in the fetus.* Pharmacology & Therapeutics, 1980. **10**(2): p. 261-281 DOI: https://doi.org/10.1016/0163-7258(80)90083-2.
29. United Nations, *Globally Harmonized System of Classification and Labelling of Chemicals (GHS Rev. 9, 2021)*. 2021.
30. Whorton, D., et al., *Infertility in male pesticide workers.* Lancet, 1977. **2**(8051): p. 1259-61 DOI: 10.1016/s0140-6736(77)92665-4.
31. Vargesson, N., *Thalidomide-induced teratogenesis: History and mechanisms.* Birth Defects Research Part C: Embryo Today: Reviews, 2015. **105**(2): p. 140-156 DOI: https://doi.org/10.1002/bdrc.21096.
32. European Commission, *Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC.* Official Journal of the European Union, 2009. **L 309/1**: p. 1-50.
33. European Commission, *Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products Text with EEA relevance.* Official Journal of the European Union, 2012. **L 167/1**: p. 1-123.
34. European Commission, *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.* Official Journal of the European Union, 2006. **L396/1**: p. 1-849.
35. OECD, *Test No. 414: Prenatal Developmental Toxicity Study*, in *OECD Guidelines for the Testing of Chemicals, Section 4*. 2018, OECD Publishing: Paris.
36. United States Environmental Protection Agency (EPA), *OPPTS 870.3700 Prenatal Developmental Toxicity Study*. Health Effects Test Guidelines. 1998: EPA.
37. OECD, *Test No. 416: Two-Generation Reproduction Toxicity*. OECD Guidelines for the Testing of Chemicals, Section 4. 2001, Paris: OECD Publishing.
38. United States Environmental Protection Agency (EPA), *OPPTS 870.3800 Reproduction and Fertility Effects*. Health Effects Test Guidelines. 1998: EPA.
39. OECD, *Test No. 421: Reproduction/Developmental Toxicity Screening Test*. OECD Guidelines for the Testing of Chemicals, Section 4. 2016, Paris: OECD Publishing.
40. United States Environmental Protection Agency (EPA), *OPPTS 870.3550 Reproduction/Developmental Toxicity Screening Test*. Health Effects Test Guidelines. 2000: EPA.
41. OECD, *Test No. 422: Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test*. OECD Guidelines for the Testing of Chemicals, Section 4. 2016, Paris: OECD Publishing.
42. United States Environmental Protection Agency (EPA), *OPPTS 870.3650 Combined Repeated Dose Toxicity with the Reproduction/Developmental Toxicity Screening Test*. Health Effects Test Guidelines. 2000: EPA.
43. OECD, *Test No. 426: Developmental Neurotoxicity Study*. OECD Guidelines for the Testing of Chemicals, Section 4. 2007, Paris: OECD Publishing.
44. United States Environmental Protection Agency (EPA), *OPPTS 870.6300 - Developmental Neurotoxicity Study*. Health Effects Test Guidelines. 1998: EPA.
45. OECD, *Test No. 443: Extended One-Generation Reproductive Toxicity Study*. OECD Guidelines for the Testing of Chemicals, Section 4. 2018, Paris: OECD Publishing.
46. DeSesso, J.M. and A.R. Scialli, *Bone development in laboratory mammals used in developmental toxicity studies.* Birth Defects Research, 2018. **110**(15): p. 1157-1187 DOI: https://doi.org/10.1002/bdr2.1350.
47. Woutersen, M., et al., *Regulating human safety: How dose selection in toxicity studies impacts human health hazard assessment and subsequent risk management options.* Regulatory Toxicology and Pharmacology, 2020. **114**: p. 104660 DOI: https://doi.org/10.1016/j.yrtph.2020.104660.

48. Buschmann, J., *The OECD Guidelines for the Testing of Chemicals and Pesticides*, in *Teratogenicity Testing: Methods and Protocols*, P.C. Barrow, Editor. 2013, Humana Press: Totowa, NJ. p. 37-56.
49. Sewell, F., et al., *Recommendations on dose level selection for repeat dose toxicity studies.* Archives of Toxicology, 2022. **96**(7): p. 1921-1934 DOI: 10.1007/s00204-022-03293-3.
50. Makris, S.L., et al., *Terminology of developmental abnormalities in common laboratory mammals (version 2).* Reproductive Toxicology, 2009. **28**(3): p. 371-434 DOI: https://doi.org/10.1016/j.reprotox.2009.06.010.
51. Chahoud, I., et al., *Classification terms in developmental toxicology: need for harmonisation 11Report of the Second Workshop on the Terminology in Developmental Toxicology Berlin, 27–28 August 1998.* Reproductive Toxicology, 1999. **13**(1): p. 77-82 DOI: https://doi.org/10.1016/S0890-6238(98)00060-4.
52. Solecki, R., et al., *Harmonisation of rat fetal skeletal terminology and classification. report of the third workshop on the terminology in developmental toxicology: Berlin, 14–16 September 2000.* Reproductive Toxicology, 2001. **15**(6): p. 713-721 DOI: https://doi.org/10.1016/S0890-6238(01)00179-4.
53. Marx-Stoelting, P., et al., *25th anniversary of the Berlin workshop on developmental toxicology: DevTox database update, challenges in risk assessment of developmental neurotoxicity and alternative methodologies in bone development and growth.* Reproductive Toxicology, 2021. **100**: p. 155-162 DOI: https://doi.org/10.1016/j.reprotox.2020.11.003.
54. Beyer, B.K., et al., *ILSI/HESI maternal toxicity workshop summary: maternal toxicity and its impact on study design and data interpretation.* Birth Defects Research Part B: Developmental and Reproductive Toxicology, 2011. **92**(1): p. 36-51 DOI: https://doi.org/10.1002/bdrb.20281.
55. Harris, C. and J.M. Rogers, *5.06 - Maternally Mediated Developmental Toxicity☆*, in *Comprehensive Toxicology (Third Edition)*, C.A. McQueen, Editor. 2018, Elsevier: Oxford. p. 86-99.
56. Lankas, G.R., M.A. Cukierski, and L.D. Wise, *The role of maternal toxicity in lovastatin-induced developmental toxicity.* Birth Defects Research Part B: Developmental and Reproductive Toxicology, 2004. **71**(3): p. 111-123 DOI: https://doi.org/10.1002/bdrb.20005.
57. Tyl, R.W., *Commentary on the role of maternal toxicity on developmental toxicity.* Birth Defects Res B Dev Reprod Toxicol, 2012. **95**(3): p. 262-6 DOI: 10.1002/bdrb.21015.
58. Danielsson, B.R., *Maternal Toxicity*, in *Teratogenicity Testing: Methods and Protocols*, P.C. Barrow, Editor. 2013, Humana Press: Totowa, NJ. p. 311-325.
59. OECD, *Guidance Document on Mammalian Reproductive Toxicity Testing and Assessment*. OECD Series on Testing and Assessment. Vol. No. 43. 2008, Paris: OECD Publishing.
60. Kluxen, F.M., et al., *Using historical control data in bioassays for regulatory toxicology.* Regulatory Toxicology and Pharmacology, 2021. **125**: p. 105024 DOI: https://doi.org/10.1016/j.yrtph.2021.105024.
61. European Commission, *Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market.* Official Journal of the European Union, 2013. **L 93/1**: p. 1–84.
62. Mylchreest, E. and S.B. Harris, *Historical Control Data in Reproductive and Developmental Toxicity Studies*, in *Teratogenicity Testing: Methods and Protocols*, P.C. Barrow, Editor. 2013, Humana Press: Totowa, NJ. p. 275-294.
63. Russel WMS., B.R., *The principles of human experimental technique*. 1959, London: Methuen & Co Ltd. of human experimental technique.
64. European Commission; Joint Research Centre (JRC), *EURL ECVAM dataset on alternative methods to animal experimentation (DB-ALM)*. 2019: European Commission, Joint Research Centre (JRC) [Dataset].
65. Selderslaghs, I.W., et al., *Development of a screening assay to identify teratogenic and embryotoxic chemicals using the zebrafish embryo.* Reprod Toxicol, 2009. **28**(3): p. 308-20 DOI: 10.1016/j.reprotox.2009.05.004.
66. Bantle, J.A., D.J. Fort, and B.L. James, *Identification of developmental toxicants using the Frog Embryo Teratogenesis Assay-Xenopus (FETAX).* Hydrobiologia, 1989. **188**(1): p. 577-585 DOI: 10.1007/BF00027826.
67. Piersma, A.H., et al., *Validation of the postimplantation rat whole-embryo culture test in the international ECVAM validation study on three in vitro embryotoxicity tests.* Altern Lab Anim, 2004. **32**(3): p. 275-307 DOI: 10.1177/026119290403200307.
68. Seiler, A.E.M. and H. Spielmann, *The validated embryonic stem cell test to predict embryotoxicity in vitro.* Nature Protocols, 2011. **6**(7): p. 961-978 DOI: 10.1038/nprot.2011.348.

69. Jelínek, R., *The Chick Embryotoxicity Screening Test (CHEST)*, in *Methods in Prenatal Toxicology*, H.-J.M. D. Neubert, T.E. Kwasigroch, Editor. 1977, Thieme: Stuttgart.
70. Valerio, L.G., *In silico toxicology for the pharmaceutical sciences.* Toxicology and Applied Pharmacology, 2009. **241**(3): p. 356-370 DOI: https://doi.org/10.1016/j.taap.2009.08.022.
71. OECD, *Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals.* OECD Papers, 2006. **6**(11): p. 1-79 DOI: DOI: 10.1787/oecd_papers-v6-ar.
72. Roy, K., S. Kar, and R.N. Das, *QSAR/QSPR Modeling: Introduction*, in *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*, K. Roy, S. Kar, and R.N. Das, Editors. 2015, Springer International Publishing: Cham. p. 1-36.
73. Hansch, C. and T. Fujita, *p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure.* Journal of the American Chemical Society, 1964. **86**(8): p. 1616-1626 DOI: 10.1021/ja01062a035.
74. Danishuddin and A.U. Khan, *Descriptors and their selection methods in QSAR analysis: paradigm for drug design.* Drug Discovery Today, 2016. **21**(8): p. 1291-1302 DOI: https://doi.org/10.1016/j.drudis.2016.06.013.
75. Cereto-Massagué, A., et al., *Molecular fingerprint similarity search in virtual screening.* Methods, 2015. **71**: p. 58-63 DOI: https://doi.org/10.1016/j.ymeth.2014.08.005.
76. Bajusz, D., A. Rácz, and K. Héberger, *3.14 - Chemical Data Formats, Fingerprints, and Other Molecular Descriptions for Database Analysis and Searching*, in *Comprehensive Medicinal Chemistry III*, S. Chackalamannil, D. Rotella, and S.E. Ward, Editors. 2017, Elsevier: Oxford. p. 329-378.
77. Rogers, D. and M. Hahn, *Extended-Connectivity Fingerprints.* Journal of Chemical Information and Modeling, 2010. **50**(5): p. 742-754 DOI: 10.1021/ci100050t.
78. Weaver, S. and M.P. Gleeson, *The importance of the domain of applicability in QSAR modeling.* Journal of Molecular Graphics and Modelling, 2008. **26**(8): p. 1315-1326 DOI: https://doi.org/10.1016/j.jmgm.2008.01.002.
79. Yang, H., et al., *Computational Approaches to Identify Structural Alerts and Their Applications in Environmental Toxicology and Drug Discovery.* Chemical Research in Toxicology, 2020. **33**(6): p. 1312-1322 DOI: 10.1021/acs.chemrestox.0c00006.
80. Venkatapathy, R. and N.C.Y. Wang, *Developmental Toxicity Prediction*, in *Computational Toxicology: Volume II*, B. Reisfeld and A.N. Mayeno, Editors. 2013, Humana Press: Totowa, NJ. p. 305-340.
81. Raies, A.B. and V.B. Bajic, *In silico toxicology: computational methods for the prediction of chemical toxicity.* Wiley Interdiscip Rev Comput Mol Sci, 2016. **6**(2): p. 147-172 DOI: 10.1002/wcms.1240.
82. Cronin, M.T.D. and A.P. Worth, *(Q)SARs for Predicting Effects Relating to Reproductive Toxicity.* QSAR & Combinatorial Science, 2008. **27**(1): p. 91-100 DOI: https://doi.org/10.1002/qsar.200710118.
83. Gupta, P.K., *Chapter 9 - Principles and basic concepts of toxicokinetics*, in *Fundamentals of Toxicology*, P.K. Gupta, Editor. 2016, Academic Press. p. 87-107.
84. Gupta, P.K., *Chapter 7 - Absorption, distribution, and excretion of toxicants*, in *Fundamentals of Toxicology*, P.K. Gupta, Editor. 2016, Academic Press. p. 55-72.
85. Lehman-McKeeman, L.D., *Absorption, Distribution, and Excretion of Toxicants*, in *Casarett and Doull's Toxicology: The Basic Science of Poisons, Eighth Edition*. 2012, McGraw-Hill Education: New York.
86. Stillwell, W., *Membrane Transport.* An Introduction to Biological Membranes, 2016: p. 423-451 DOI: 10.1016/B978-0-444-63772-7.00019-1.
87. Eisenbrand, G., M. Metzler, and F.J. Hennecke, *Toxikologie für Naturwissenschaftler und Mediziner: Stoffe, Mechanismen, Prüfverfahren*. 2005: Wiley.
88. Liesivuori, J. and H. Savolainen, *Methanol and formic acid toxicity: biochemical mechanisms.* Pharmacol Toxicol, 1991. **69**(3): p. 157-63 DOI: 10.1111/j.1600-0773.1991.tb01290.x.
89. Taft, D.R., *Chapter 9 - Drug Excretion*, in *Pharmacology*, M. Hacker, W. Messer, and K. Bachmann, Editors. 2009, Academic Press: San Diego. p. 175-199.
90. Boobis, A.R., et al., *IPCS Framework for Analyzing the Relevance of a Cancer Mode of Action for Humans.* Critical Reviews in Toxicology, 2006. **36**(10): p. 781-792 DOI: 10.1080/10408440600977677.
91. Edwards, S.W., et al., *Adverse Outcome Pathways—Organizing Toxicological Information to Improve Decision Making.* Journal of Pharmacology and Experimental Therapeutics, 2016. **356**(1): p. 170-181 DOI: 10.1124/jpet.115.228239.

92. OECD, *Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA)*, in *OECD Series on Testing and Assessment*. 2017, OECD Publishing: Paris.
93. OECD, *Users' Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways*, in *OECD Series on Adverse Outcome Pathways*. 2018, OECD Publishing: Paris.
94. OECD, *Guidance Document for the scientific review of Adverse Outcome Pathways*. 2021, OECD Publishing: Paris.
95. Ito, T. and H. Handa, *Deciphering the mystery of thalidomide teratogenicity.* Congenital Anomalies, 2012. **52**(1): p. 1-7 DOI: https://doi.org/10.1111/j.1741-4520.2011.00351.x.
96. Gao, S., et al., *Recent advances in the molecular mechanism of thalidomide teratogenicity.* Biomedicine & Pharmacotherapy, 2020. **127**: p. 110114 DOI: https://doi.org/10.1016/j.biopha.2020.110114.
97. Drozdzik, M., et al., *Protein abundance of pharmacokinetically relevant transporter proteins and metabolizing enzymes along the human intestine and in the liver: A comparative intra-subject study.* Drug Metabolism and Pharmacokinetics, 2018. **33**(1, Supplement): p. S76-S77 DOI: https://doi.org/10.1016/j.dmpk.2017.11.252.
98. Gao, B., et al., *Differential expression of bile salt and organic anion transporters in developing rat liver.* J Hepatol, 2004. **41**(2): p. 201-8 DOI: 10.1016/j.jhep.2004.04.029.
99. Hoffmaster, K.A., et al., *P-glycoprotein Expression, Localization, and Function in Sandwich-Cultured Primary Rat and Human Hepatocytes: Relevance to the Hepatobiliary Disposition of a Model Opioid Peptide.* Pharmaceutical Reasearch, 2004. **21**(7).
100. Tomer, G., et al., *Differential developmental regulation of rat liver canalicular membrane transporters Bsep and Mrp2.* Pediatr Res, 2003. **53**(2): p. 288-94 DOI: 10.1203/01.PDR.0000047509.54253.01.
101. Gu, X. and J.E. Manautou, *Regulation of hepatic ABCC transporters by xenobiotics and in disease states.* Drug Metab Rev, 2010. **42**(3): p. 482-538 DOI: 10.3109/03602531003654915.
102. Klaassen, C.D. and L.M. Aleksunes, *Xenobiotic, bile acid, and cholesterol transporters: function and regulation.* Pharmacol Rev, 2010. **62**(1): p. 1-96 DOI: 10.1124/pr.109.002014.
103. Murray, M. and F. Zhou, *Trafficking and other regulatory mechanisms for organic anion transporting polypeptides and organic anion transporters that modulate cellular drug and xenobiotic influx and that are dysregulated in disease.* Br J Pharmacol, 2017. **174**(13): p. 1908-1924 DOI: 10.1111/bph.13785.
104. Meier, Y., et al., *Interindividual variability of canalicular ATP-binding-cassette (ABC)–transporter expression in human liver.* Hepatology, 2006. **44**(1): p. 62-74 DOI: https://doi.org/10.1002/hep.21214.
105. Wolters, H., et al., *Isolation and characterization of canalicular and basolateral plasma membrane fractions from human liver.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 1991. **1069**(1): p. 61-69 DOI: https://doi.org/10.1016/0005-2736(91)90104-G.
106. Shitara, Y., et al., *Clinical significance of organic anion transporting polypeptides (OATPs) in drug disposition: their roles in hepatic clearance and intestinal absorption.* Biopharm Drug Dispos, 2013. **34**(1): p. 45-78 DOI: 10.1002/bdd.1823.
107. Wang, D., *Current Research Method in Transporter Study*, in *Drug Transporters in Drug Disposition, Effects and Toxicity*, X. Liu and G. Pan, Editors. 2019, Springer Singapore: Singapore. p. 203-240.
108. Blazquez, A.G., et al., *Characterization of the role of ABCG2 as a bile acid transporter in liver and placenta.* Mol Pharmacol, 2012. **81**(2): p. 273-83 DOI: 10.1124/mol.111.075143.
109. Cvetkovic, M., et al., *OATP and P-Glycoprotein Transporters Mediate the Cellular Uptake and Excretion of Fexofenadine.* Drug Metabolism and Disposition, 1999. **27**(8): p. 866-871.
110. De Bruyn, T., et al., *Sandwich-cultured hepatocytes: utility for in vitro exploration of hepatobiliary drug disposition and drug-induced hepatotoxicity.* Expert Opin Drug Metab Toxicol, 2013. **9**(5): p. 589-616 DOI: 10.1517/17425255.2013.773973.
111. Fattah, S., P. Augustijns, and P. Annaert, *Age-dependent activity of the uptake transporters Ntcp and Oatp1b2 in male rat hepatocytes: from birth till adulthood.* Drug Metab Dispos, 2015. **43**(1): p. 1-8 DOI: 10.1124/dmd.114.059212.
112. Maeda, A., et al., *Drug interaction between celecoxib and methotrexate in organic anion transporter 3–transfected renal cells and in rats in vivo.* European Journal of Pharmacology, 2010. **640**(1): p. 168-171 DOI: https://doi.org/10.1016/j.ejphar.2010.04.051.
113. Mölsä, M., et al., *Functional role of P-glycoprotein in the human blood-placental barrier.* Clinical Pharmacology & Therapeutics, 2005. **78**(2): p. 123-131 DOI: 10.1016/j.clpt.2005.04.014.

114. Klaassen, C.D. and H. Lu, *Xenobiotic Transporters: Ascribing Function from Gene Knockout and Mutation Studies.* Toxicological Sciences, 2007. **101**(2): p. 186-196 DOI: 10.1093/toxsci/kfm214.
115. Aller, S.G., et al., *Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding.* Science, 2009. **323**(5922): p. 1718-22 DOI: 10.1126/science.1168750.
116. European Commission, *Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes.* Official Journal of the European Union, 2010. **L 276/33**: p. 33-79.
117. STAFFORD, D.W., *The vitamin K cycle.* Journal of Thrombosis and Haemostasis, 2005. **3**(8): p. 1873-1878 DOI: https://doi.org/10.1111/j.1538-7836.2005.01419.x.
118. Mladěnka, P., et al., *Vitamin K - sources, physiological role, kinetics, deficiency, detection, therapeutic use, and toxicity.* Nutrition reviews, 2022. **80**(4): p. 677-698 DOI: 10.1093/nutrit/nuab061.
119. Hirota, Y. and Y. Suhara, *New Aspects of Vitamin K Research with Synthetic Ligands: Transcriptional Activity via SXR and Neural Differentiation Activity.* International Journal of Molecular Sciences, 2019. **20**(12): p. 3006.
120. Stahmann, M.A., C.F. Huebner, and K.P. Link, *STUDIES ON THE HEMORRHAGIC SWEET CLOVER DISEASE: V. IDENTIFICATION AND SYNTHESIS OF THE HEMORRHAGIC AGENT.* Journal of Biological Chemistry, 1941. **138**(2): p. 513-527 DOI: https://doi.org/10.1016/S0021-9258(18)51377-6.
121. Watt, B.E., et al., *Anticoagulant Rodenticides.* Toxicological Reviews, 2005. **24**(4): p. 259-269 DOI: 10.2165/00139709-200524040-00005.
122. Kim, S., et al., *PubChem in 2021: new data content and improved web interfaces.* Nucleic acids research, 2021. **49**(D1): p. D1388-D1395 DOI: 10.1093/nar/gkaa971.
123. Hall, J.G., R.M. Pauli, and K.M. Wilson, *Maternal and fetal sequelae of anticoagulation during pregnancy.* The American Journal of Medicine, 1980. **68**(1): p. 122-140 DOI: 10.1016/0002-9343(80)90181-3.
124. Yurdakök, M., *Fetal and neonatal effects of anticoagulants used in pregnancy: a review.* The Turkish journal of pediatrics, 2012. **54 3**: p. 207-15.
125. Schaefer, C., et al., *Vitamin K antagonists and pregnancy outcome.* Thromb Haemost, 2006. **95**(06): p. 949-957.
126. MENON, R.K., et al., *Impaired Carboxylation of Osteocalcin in Warfarin-Treated Patients.* The Journal of Clinical Endocrinology & Metabolism, 1987. **64**(1): p. 59-61 DOI: 10.1210/jcem-64-1-59.
127. Raghav, S. and D. Reutens, *Neurological sequelae of intrauterine warfarin exposure.* Journal of Clinical Neuroscience, 2007. **14**(2): p. 99-103 DOI: https://doi.org/10.1016/j.jocn.2006.03.031.
128. Breckenridge, A.M., et al., *A study of the relationship between the pharmacokinetics and the pharmacodynamics of the 4-hydroxycoumarin anticoagulants warfarin, difenacoum and brodifacoum in the rabbit.* British journal of pharmacology, 1985. **84**(1): p. 81-91.
129. Mosterd, J.J. and H.H. Thijssen, *The long-term effects of the rodenticide, brodifacoum, on blood coagulation and vitamin K metabolism in rats.* British journal of pharmacology, 1991. **104**(2): p. 531-535 DOI: 10.1111/j.1476-5381.1991.tb12463.x.
130. Huckle, K.R., D.H. Hutson, and P.A. Warburton, *Elimination and Accumulation of the Rodenticide Flocoumafen in Rats Following Repeated Oral Administration.* Xenobiotica, 1988. **18**(12): p. 1465-1479 DOI: 10.3109/00498258809042269.
131. VANDENBROUCKE, V., et al., *Pharmacokinetics of eight anticoagulant rodenticides in mice after single oral administration.* Journal of Veterinary Pharmacology and Therapeutics, 2008. **31**(5): p. 437-445 DOI: https://doi.org/10.1111/j.1365-2885.2008.00979.x.
132. Bajoria, R., S. Sooranna, and R. Chatterjee, *Effect of lipid composition of cationic SUV liposomes on materno-fetal transfer of warfarin across the perfused human term placenta.* Placenta, 2013. **34**(12): p. 1216-1222 DOI: https://doi.org/10.1016/j.placenta.2013.10.005.
133. Chetot, T., et al., *Differences in teratogenicity of some vitamin K antagonist substances used as human therapeutic or rodenticide are due to major differences in their fate after an oral administration.* Toxicology Letters, 2020. **333**: p. 71-79 DOI: https://doi.org/10.1016/j.toxlet.2020.07.034.
134. Chetot, T., et al., *Vitamin K antagonist rodenticides display different teratogenic activity.* Reproductive Toxicology, 2020. **93**: p. 131-136 DOI: https://doi.org/10.1016/j.reprotox.2020.02.003.
135. Myden, A., E. Hill, and A. Fowkes, *Using adverse outcome pathways to contextualise (Q)SAR predictions for reproductive toxicity – A case study with aromatase inhibition.* Reproductive Toxicology, 2022. **108**: p. 43-55 DOI: https://doi.org/10.1016/j.reprotox.2022.01.004.

136. Lasseur, R., et al., *Comparison of the inhibition effect of different anticoagulants on vitamin K epoxide reductase activity from warfarin-susceptible and resistant rat.* Pesticide Biochemistry and Physiology, 2007. **88**(2): p. 203-208 DOI: https://doi.org/10.1016/j.pestbp.2006.11.005.
137. Tie, J.K. and D.W. Stafford, *Chapter Fourteen - Functional Study of the Vitamin K Cycle Enzymes in Live Cells*, in *Methods in Enzymology*, M.H. Gelb, Editor. 2017, Academic Press. p. 349-394.
138. Aengenheister, L., et al., *An advanced human in vitro co-culture model for translocation studies across the placental barrier.* Scientific Reports, 2018. **8**(1): p. 5388 DOI: 10.1038/s41598-018-23410-6.
139. Levkovitz, R., et al., *In vitro simulation of placental transport: Part I. Biological model of the placental barrier.* Placenta, 2013. **34**(8): p. 699-707 DOI: https://doi.org/10.1016/j.placenta.2013.03.014.
140. Litsa, E.E., P. Das, and L.E. Kavraki, *Prediction of drug metabolites using neural machine translation.* Chemical Science, 2020. **11**(47): p. 12777-12788 DOI: 10.1039/D0SC02639E.
141. Wang, D., et al., *Deep Learning Based Drug Metabolites Prediction.* Front Pharmacol, 2019. **10**: p. 1586 DOI: 10.3389/fphar.2019.01586.
142. Gupta, R.C., *Introduction*, in *Reproductive and Developmental Toxicology (Third Edition)*, R.C. Gupta, Editor. 2022, Academic Press. p. xix-xxi.
143. Juliano, R.L. and V. Ling, *A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 1976. **455**(1): p. 152-162 DOI: https://doi.org/10.1016/0005-2736(76)90160-7.
144. Liu, X., *ABC Family Transporters*, in *Drug Transporters in Drug Disposition, Effects and Toxicity*, X. Liu and G. Pan, Editors. 2019, Springer Singapore: Singapore. p. 13-100.
145. Liu, X., *SLC Family Transporters*, in *Drug Transporters in Drug Disposition, Effects and Toxicity*, X. Liu and G. Pan, Editors. 2019, Springer Singapore: Singapore. p. 101-202.

6 Appendix

6.1 List of Abbreviations

ABC	ATP-binding cassette
ACC	Accuracy
ADME	Absorption, Distribution, Metabolism and Excretion
AOP	Adverse outcome pathways
BA	Balanced accuracy
BCRP	Breast cancer resistance protein
BSEP	Bile salt export pump
CHEST	Chicken Embryotoxicity Test
DB-ALM	Database on Alternative Methods
DBCD	1,2-Dibromo-3-chloropropane
ECHA	European Chemicals Agency
EFD	Embryo-Fetal Developmental
EFSA	European Food Safety Authority
EPA	Environmental Protection Agency, US
EST	Embryonic stem cell test
EU	European Union
EURL ECVAM	European Union Reference Laboratory for alternatives to animal testing
FEED	Fertility and Early Embryonic Development
FETAX	Frog embryo teratogenesis assay Xenopus
FWS	Foetal warfarin syndrome
GD	Gestation day
GHS	Globally Harmonised System of Classification and Labelling of Chemicals
kDa	kilodalton
KE	Key event
KER	Key event relationship
MDR	Multidrug resistance protein
MIE	Molecular initiating event
MoA	Mode of action
MOA	Mechanism of action
MRP	Multidrug resistance-related protein
NOAEL	No observed adverse effect level
OAT	Organic anion transporter
OATP	Organic anion transporting polypeptide
OECD	Organization for Economic Co-operation and Development
OPPTS	Office of Prevention, Pesticides & Toxic Substances

PBPK modelling	Physiologically based pharmacokinetic modelling
P-gp	P-glycoprotein
PND	Postnatal day
PPND	Pre- and Postnatal Developmental
qPCR	quantitative polymerase chain reaction
QSAR	Quantitative Structure–Activity Relationship
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
Reprotoxicity	Reproductive toxicity
SA	Structural alert
SEN	Sensitivity
SLC	Solute carrier
SPC	Specifity
TDAR	T-cell-dependent antibody response assay
TP	Time point
WEC	Whole embryo culture test
wpc	Weeks post conception
ZET	Zebra fish embryotoxicity test

6.2 Danksagung

An dieser Stelle möchte ich mich herzlich bei allen bedanken, die durch Ihre fachliche und/oder persönliche Unterstützung zum Gelingen dieser Arbeit beigetragen haben.

Allen voran möchte ich meinen Betreuern Thomas Hofmann und Markus Frericks für die Bereitstellung der faszinierenden Themen und die Möglichkeit, meine Dissertation bei der BASF SE in Ludwigshafen anzufertigen danken. Die Corona-Pandemie hat diese Dissertation zu einer besonderen Herausforderung werden lassen und einen Themenwechsel erforderlich gemacht. Zusammen mit euch habe ich diese schwierige Zeit gemeistert. Danke für eurer Vertrauen in mich, eure Unterstützung und die interessanten Diskussionen. Ich habe sehr viel von euch gelernt!

Bei Elke Richling möchte ich mich für die Übernahme der Betreuung dieser Dissertation bedanken.

Steffen Schneider und dem gesamten Team des Labors Reproduktionstoxikologie danke ich für ihre professionelle technische Unterstützung bei der Probengewinnung und ihre große Hilfsbereitschaft bei all meinen Fragen zum Thema Reprotox.

Michael Eichenlaub und dem gesamten Team des Labors Industrielle Biotechnologie danke ich für ihre Unterstützung bei den qPCR Experimenten und viele sehr amüsante und inspirierende Gespräche.

Steven Van Cruchten danke ich für seine fachlichen und sprachlichen Korrekturen bei der Erstellung der ersten Publikation.

Geertje Lewin und Madeleine Joel von Preclinical Science danke ich für die großartige Zusammenarbeit bei der Erstellung der zweiten Publikation.

Meinen Mitdoktoranden Pia Wilhelmi, Alina Göpfert, Sabina Ramirez Hincapie, Dunja Dimitrijevic und Andreas Weber danke ich für ihre Hilfsbereitschaft, die konstruktiven wissenschaftlichen Gespräche und weniger wissenschaftlichen Plaudereien. Unsere gemeinsamen Mittagessen waren stehts sehr unterhaltsam, lustig und motivierend.

Meiner ehemaligen Kollegin Jana Keller danke ich für ihre sehr konstruktive Kritik an dieser Arbeit. Außerdem für die tolle Übernachtungsmöglichkeit mit leckem Essen. Ich habe mich bei euch immer wie zuhause gefühlt!

Ein besonderer Dank gilt meiner Familie. Ihr habt mich immer unterstützt, an mich geglaubt und meine Launen ausgehalten. Ohne euch hätte ich das nicht geschafft!

6.3 Lebenslauf

Akademische Ausbildung

Seit 07/2019	Doktorandin im Bereich „Experimental Toxicology & Ecology“ der BASF SE, Ludwigshafen
12/2018	Masterabschluss (Note: 1,12)
2018	9-monatiges Praktikum mit Masterarbeit im Labor von Prof. Viktor Stein, TU Darmstadt
10/2016 bis 12/2018	Masterstudium im Fach „Biomolecular Engineering“ an der TU Darmstadt
09/2016	Bachelorabschluss (Note: 1,3)
2016	4-monatiges Praktikum mit Bachelorarbeit im Labor von Prof. Barbara Schnierle, Paul-Ehrlich-Institute, Langen
10/2013 bis 09/2016	Bachelorstudium im Fach „Biomolecular Engineering“ an der TU Darmstadt
06/2013	Allgemeine Hochschulreife am Gymnasium Michelstadt (Note: 1,2)

Fortbildungen

02/2022	Weiterbildungsprogramm zur Anerkennung als “Fachtoxikologe/Fachtoxikologin DGPT“ – Grundlagen der Organtoxikologie und -pathologie, Teil 1
06/2021	Weiterbildungsprogramm zur Anerkennung als “Fachtoxikologe/Fachtoxikologin DGPT“ – Grundlagen der Organtoxikologie und -pathologie, Teil 2
01/2021	Reproductive and Developmental Toxicology Course der European Teratology Society

Stipendien und Preise

2022	GT-Preis des Arbeitskreises Computational Toxicology 2022
2016	Preis der Dr.-Anton-Keller-Stiftung für sehr gute Leistungen in der B.Sc. BME-Abschlussprüfung
10/2014 bis 09/2018	Stipendiatin des Deutschlandstipendiums

Veröffentlichungen

Weyrich, A., Joel, M., Lewin, G., Hofmann, T., & Frericks, M. (2022). Review of the state of science and evaluation of currently available *in silico* prediction models for reproductive and developmental toxicity: A case study on pesticides. *Birth Defects Research*, 1– 31. https://doi.org/https://doi.org/10.1002/bdr2.2062

Weyrich, A., Joel, M., Lewin, G., Hofmann, T.& Frericks, M. (2022). Reliability of published QSAR models for the prediction of developmental and reproductive toxicity – a case study on pesticides. Digital poster session and talk presented at German Pharm-Tox Summit 2022

Weyrich, A., Frericks, M., Eichenlaub, M., Schneider, S., Hofmann, T., Van Cruchten, S., & van Ravenzwaay, B. (2022). Ontogeny of renal, hepatic, and placental expression of ATP-binding cassette and solute carrier transporters in the rat and the rabbit. *Reproductive Toxicology*, 107, 1-9. https://doi.org/10.1016/j.reprotox.2021.10.005

Weyrich, A., Frericks, M., Hofmann, T., Schneider, S., Eichenlaub, M. & van Ravenzwaay, B. (2021). Expression of xenobiotic transporters in the rat and rabbit kidney and liver during pre- and postnatal development. Digital poster session presented at EUROTOX Virtual Congress 2021

Weyrich, A., Frericks, M., Hofmann, T., Schneider, S., Eichenlaub, M. & van Ravenzwaay, B. (2021). Species specific and developmental influences on the expression of xenobiotic transporters. Digital poster session presented at German Pharm-Tox Summit 2021

Weyrich, A., Frericks, M., Hofmann, T., Schneider, S., Eichenlaub, M. & van Ravenzwaay, B. (2020). Expression of xenobiotic transporters during rat development. Poster session presented at German Pharm-Tox Summit 2020, Leipzig

Fachbereich Chemie
Technische Universität Kaiserslautern
67653 Kaiserslautern
Telefon: +49 631 205-2754
Telefax: +49 631 205-2187
www.chemie.uni-kl.de

TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN