Regulatory Toxicology and Pharmacology 123 (2021) 104931

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

New framework for a non-animal approach adequately assures the safety of cosmetic ingredients – A case study on caffeine

Dagmar Bury [a], Camilla Alexander-White [b], Harvey J. Clewell III [c], Mark Cronin [d], Bertrand Desprez [e], Ann Detroyer [f], Alina Efremenko [g], James Firman [d], Eric Hack [g], Nicola J. Hewitt [h], Gerry Kenna [e], Martina Klaric [e], Cathy Lester [h], Catherine Mahony [i], Gladys Ouedraogo [f], Alicia Paini [j], Andreas Schepky [k], on behalf of Cosmetics Europe [*]

[a] L'Oréal, Research & Innovation, 9 Rue Pierre Dreyfus, 92110, Clichy, France
[b] MKTox & Co Ltd, 36 Fairford Crescent, Downhead Park, Milton Keynes, Buckinghamshire, MK15 9AQ, UK
[c] Ramboll Health Sciences, 3107 Armand Street, Monroe, LA, 71201, USA
[d] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool, L3 AF, UK
[e] Cosmetics Europe, 40 Avenue Hermann-Debroux, 1160, Brussels, Belgium
[f] L'Oréal, Research & Innovation, 1 Avenue Eugène Schueller, Aulnay-sous-Bois, France
[g] ScitoVation, Research Triangle Park, Durham, NC, USA
[h] The Procter & Gamble Co., Cincinnati, OH, USA
[i] Procter & Gamble Technical Centres Ltd, Reading, RG2 0RX, UK
[j] European Commission Joint Research Centre, Ispra, Italy
[k] Beiersdorf, Unnastrasse 48, 20253, Hamburg, Germany

ARTICLE INFO

Handling Editor: Dr. Lesa Aylward

Keywords:
Caffeine
Read-across
New approach methodologies
Next generation risk assessment
Systemic toxicity
Physiologically-based kinetic modelling

ABSTRACT

This case study on the model substance caffeine demonstrates the viability of a 10-step read-across (RAX) framework in practice. New approach methodologies (NAM), including RAX and physiologically-based kinetic (PBK) modelling were used to assess the consumer safety of caffeine. Appropriate animal systemic toxicity data were used from the most relevant RAX analogue while assuming that no suitable animal toxicity data were available for caffeine. Based on structural similarities, three primary metabolites of the target chemical caffeine (theophylline, theobromine and paraxanthine) were selected as its most relevant analogues, to estimate a point of departure in order to support a next generation risk assessment (NGRA). On the basis of the pivotal mode of action (MOA) of caffeine and other methylxanthines, theophylline appeared to be the most potent and suitable analogue. A worst-case aggregate exposure assessment determined consumer exposure to caffeine from different sources, such as cosmetics and food/drinks. Using a PBK model to estimate human blood concentrations following exposure to caffeine, an acceptable Margin of Internal Exposure (MOIE) of 27-fold was derived on the basis of a RAX using theophylline animal data, which suggests that the NGRA approach for caffeine is sufficiently conservative to protect human health.

1. Introduction

Caffeine (IUPAC name 1,3,7-trimethylpurine-2,6-dione; CAS number 58-08-2) is a naturally occurring methylxanthine found in the cherry beans of *Rubiaceae* plants. Consumers may be exposed to caffeine through foods and beverages, food supplements, medicines and cosmetic products. According to the European Regulation (EC) 1223/2009, cosmetic products placed on the European Union (EU) market must be

Abbreviations: ADME, absorption, distribution, metabolism, excretion; CPR, Cosmetic Products Regulation; CSR, Cosmetic Safety Report; EU, European Union; IP, intraperitoneal; IV, intravenous; MOA, mode of action; MOIE, Margin of Internal Exposure; NAM, New Approach Methodologies; NGRA, Next Generation Risk Assessment; PBK, physiologically-based kinetic; POD, point of departure; PBS, phosphate-buffered saline; RAX, Read-Across; RPF, relative potency factor; SCCS, Scientific Committee on Consumer Safety; SEURAT, Safety Evaluation Ultimately Replacing Animal Testing; SMILES, Simplified Molecular Input Line Entry Specification; TTC, Threshold of Toxicological Concern.

* Corresponding author. Cosmetics Europe, Avenue Hermann Debroux 40, 1160 Brussels, Belgium.
E-mail addresses: dagmar.bury@rd.loreal.com, cosmeticseurope@cosmeticseurope.eu (D. Bury).

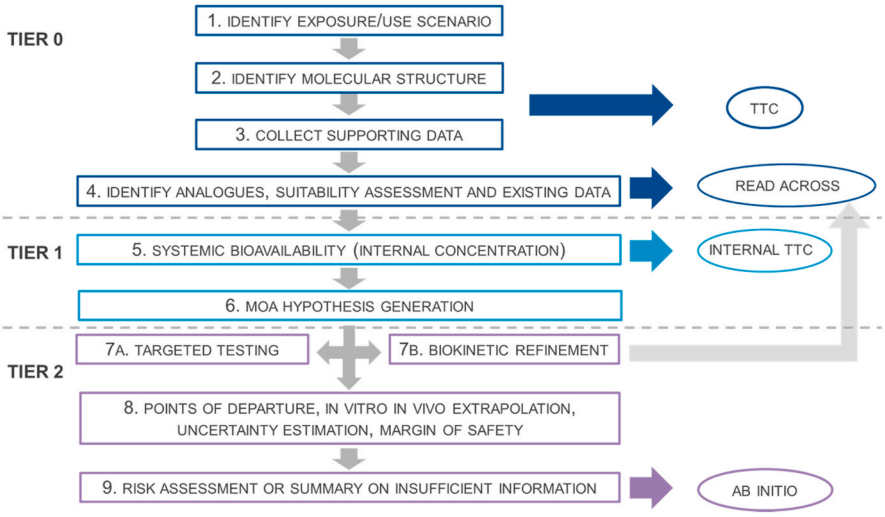
https://doi.org/10.1016/j.yrtph.2021.104931
Received 25 November 2020; Received in revised form 11 March 2021; Accepted 13 April 2021
Available online 24 April 2021

safe. Their manufacturer must ensure that they undergo an expert scientific safety assessment before they are put on the market. Caffeine is not used as a preservative, colourant or UV filter, and is therefore not listed in the annexes of the EU Cosmetic Products Regulation (CPR), but for its Cosmetic Safety Report (CSR) a safety assessment is required.

In the EU, if there are safety data gaps or a potential new human health issue arises for a chemical that is used in cosmetic products, it is not possible to use newly generated animal data to fill those gaps given the ban on animal testing for cosmetic ingredients that came into force in the EU in March 2013. Therefore, new ways had to be found to assure the safety of cosmetics ingredients in the absence of animal testing data. In an accompanying publication (Alexander-White et al., 2021), an overarching 10-step framework is described according to the recommendations of the SEURAT-1 project, using read-across (RAX) and physiologically-based kinetic (PBK) modelling in order to perform a next generation risk assessment (NGRA) to determine the safety of cosmetic ingredients in a regulatory decision-making context. To this end, in this case study we set out to validate whether the application of this tiered multiple step framework is a realistic approach to perform a pragmatic and sufficiently conservative risk assessment without the use of animal data for the target chemical caffeine.

Caffeine is a well-known and data-rich compound for which various animal and human data are available covering its pharmacological and toxicological effects. Since caffeine has been consumed and used in products for centuries, a "history of safe use" rationale has been used to support its safety (CIR 2019; EFSA 2015). In this case study, however,

a)



b)

Tier 0

Step 1: Identify exposure/use scenarios for target chemical

Step 2: Identify molecular structure of target chemical

Step 3: Collate supporting data on target chemical and define data gap(s)

Step 4: Analogue(s) a) Identify, b) collate existing data, c) determine similarity hypothesis

End Tier 0 → Potential to move to Steps 8-10 if data are sufficient

Tier1

Step 5: Systemic bioavailability/ADME of target chemical and analogues

Step 6: Supporting a similar Mode of Action (MOA) hypothesis

End Tier 1 → Potential to move to Steps 8-10 if data are sufficient

Tier 2

Step 7: a) Perform targeted testing using New Approach Methodology (NAM) assays to strengthen hypotheses and/or b) kinetic refinements of target chemical and analogues

Step 8: Performing a read-across (RAX) to derive a point of departure (POD)

Step 9: Performing a margin of safety (MOS) evaluation

Step 10 Assessing the level of confidence in the safety assessment approach

Fig. 1. (a) The SEURAT-1 workflow to perform a risk assessment without new animal data (Berggren et al., 2017) and (b) the principles of the SEURAT-1 workflow in a 10-step framework applying read-across (RAX) and New Approach Methodologies (NAM) in a next generation risk assessment (NGRA) (Alexander-White et al., 2021).

we assumed that no animal systemic toxicity data were available for caffeine. Our hypothesis was that in this situation, RAX can be used, drawing upon animal systemic toxicity data from one or more suitable analogue(s) of caffeine, in order to define a point of departure (POD). This POD together with exposure data will then be used to support a NGRA.

At the start of the process, the search for suitable RAX analogues, using a range of data sources, is open to whatever chemical structures, physicochemical properties, metabolism and biological/toxicological data arise. Finally, the selection of the RAX analogues has to be justified on the basis of the available information (Alexander-White et al., 2021). This paper illustrates the application of the proposed 10-step framework in order to arrive at a risk estimate for a conservative human safety decision on caffeine.

2. Applying the 10-step RAX framework in a NGRA for caffeine

This NGRA approach follows the recommendations of the SEURAT-1 project (Berggren et al., 2017), resulting in the tiered 10-step framework (Alexander-White et al., 2021) that is applied in this case study for caffeine, as shown in Fig. 1. Starting with the problem formulation we go step-by-step through the framework.

For all NGRA it is important to begin with a clear problem formulation. In this case, human safety of the target chemical caffeine has to be assured despite the (hypothetical) lack of animal systemic toxicity data. It has to be decided what exposure represents an acceptable 'safe' concentration, without the option for new animal testing as products containing caffeine are marketed in Europe. To begin with, data are collected on the use and exposure to caffeine to estimate an external dose metric.

3. Tier 0 - steps 1 to 4 of RAX

Tier 0 of the framework does not involve any new data generation, but comprises exposure estimation, data searching and analogue identification for a RAX.

Step 1: Identify Use/Exposure scenario for target chemical

Exposure to caffeine may occur from different sources, particularly through food and beverages, but also from cosmetic products. The initial aim for this case study is to develop an aggregate external exposure dose metric for caffeine from these sources.

Using the Tier 0 deterministic cosmetic product exposure assessment, as outlined in the SCCS Notes of Guidance (SCCS, 2021), a worst-case aggregate exposure estimate from dermal use of cosmetic products - including shower gel, shampoo, hair styling products, body lotion, face cream, hand cream, liquid foundation, lipstick, and deodorant/antiperspirant - resulted in a potential maximum aggregate external exposure of 230.96 mg/kg bw/day including the use of retention factors representing the fraction on the skin available for uptake depending on the respective cosmetic product (Table 1). Assuming that caffeine is included maximally at 2% in each of the 10 product types, then aggregate external exposure to caffeine via the dermal route is estimated at 4.6 mg/kg/day (230.96 mg/kg bw/day x 0.02). This is a very conservative exposure scenario as it assumes that all cosmetic products contain caffeine and are used by all consumers at a high amount per use, simultaneously at a high frequency per day.

Caffeine exposure from other sources may include oral intake from coffee, tea, energy drinks, cola and chocolate (EFSA, 2015). The maximum 95th percentile of caffeine intake from all food/drink sources for all days is estimated to be 648 mg/person/d (10.8 mg/kg bw) for adults (18 to <65 yr) or 786 mg/person/d (13.1 mg/kg bw) for the elderly (≥65 yr). The latter was used as the worst case contributing to the assessment of the aggregate exposure to caffeine from different sources.

As explained by Alexander-White et al. (2021), in principle it is possible to exit the framework at the end of Tier 0 for a chemical with defined exposure scenario, if a threshold of toxicological concern (TTC) approach applies at this point. However, with a worst-case dermal exposure metric for caffeine in cosmetics products at 4.6 mg/kg/day and a high-end oral exposure of 13.1 mg/kg/day and keeping in mind that caffeine has a chemical structure corresponding to Cramer class I, a TTC approach is not possible as its exposure is higher than the levels allowed for a Cramer class I substance, i.e. 30 μg/kg/day. In addition, exposure refinement using probabilistic population-based approaches (https://www.cremeglobal.com/; and as exemplified in Tozer et al., 2015, for zinc pyrithione and in Tozer et al., 2019, for vitamin A exposure), was not expected to lead to a consumer exposure low enough to allow the application of the TTC approach. Hence, in order to cover the systemic toxicity data gap RAX was performed, using structurally (and potentially biologically) similar analogues with suitable data to support a reasonable estimation of a POD for caffeine.

Step 2: Identify molecular structure of target chemical

Table 1
Daily exposure estimates for different cosmetic product categories in Europe which may contain caffeine, calculated by multiplying daily amounts and retention factors (SCCS, 2021).

Product Type	Estimated daily amount applied (g/d)	Relative daily amount applied (mg/kg bw/d)[a]	Retention factor[b]	Calculated daily exposure (g/d)	Calculated relative daily exposure (mg/kg bw/d)[a]
Bathing, showering					
Shower gel	18.67	279.20	0.01	0.19	2.79
Hair care					
Shampoo	10.46	150.49	0.01	0.11	1.51
Hair styling products	4.00	57.40	0.10	0.40	5.74
Skin care					
Body lotion	7.82	123.20	1.00	7.82	123.20
Face cream	1.54	24.14	1.00	1.54	24.14
Hand cream	2.16	32.70	1.00	2.16	32.70
Make-up					
Liquid foundation	0.51	7.90	1.00	0.51	7.90
Lipstick, lip salve	0.057	0.90	1.00	0.057	0.90
Deodorant					
Deodorant non- spray	1.50	22.08	1.00	1.50	22.08
Deodorant spray	0.69	10.00	1.00	0.69	10.00
Aggregate exposure					230.96

[a] The specific body weight of the persons involved in the study is used and not the default value of 60 kg.
[b] The retention factor was introduced to take into account the fraction of the cosmetic product which is retained on the skin such as 1 for leave-on cosmetics (e.g. creams, body lotion) and 0.01–0.1 for rinse-off cosmetics (e.g. shower gel, shampoo).

The target chemical caffeine (1,3,7-trimethylpurine-2,6-dione) appears as odorless white powder or white glistening needles with bitter taste. It consists of a xanthine scaffold with three N-methyl groups as shown in Fig. 2. For the purpose of this case study, it was assumed that caffeine was a pure chemical and did not contain any impurities.

Step 3: Collate supporting data for caffeine and define data gap(s)

All attempts were made to collate data for caffeine on physicochemical properties (see Table 2), absorption, distribution, metabolism and excretion (ADME), and toxicity endpoints. A literature and database search was performed using the major authoritative worldwide sources of information such as ChemSpider, PubMed, ECHA (REACH), NTP, OECD, CIR etc. (Alexander-White et al., 2021).

According to the Caco-2 cell monolayer model caffeine is supposed to be class I of the Biopharmaceutics Classification System (BCS), which means it is a well absorbed compound of high solubility and high permeability (Smetanova et al., 2009). Model-based correlation studies of single dose exposures of human volunteers to caffeine indicated that *in vitro*-derived liver clearance of caffeine is about 10-fold lower than fitted *in vivo* clearance, even when recombinant enzymes were used (Gajewska et al., 2015). Results are also dependent on the incubation duration. Shibata et al. (2002) used 10 μM caffeine and a 2-h incubation; however, the derived intrinsic clearance (Cl_{int}) value was lower than that of other groups who used longer durations (Berthou et al., 1988), suggesting that this incubation may not have detected sufficient parent compound depletion to make an accurate calculation of the CL_{int}. Berthou et al. (1989) calculated a half-life (t1/2) of hepatic elimination for a normal average drug intake of 1400 μM (about 270 mg of caffeine) to be 4.5 h, assuming that the caffeine metabolism is linear with time in hepatocyte cultures. This value is in agreement with the t1/2 determined *in vivo* (Bonati and Garattini, 1984).

It was demonstrated that the biotransformation of caffeine *in vitro* is comparable in human liver slices, microsomes and hepatocyte cultures. In general, when unlabelled compound was added to cellular incubations at low concentrations (up to 200 μM) or the duration of incubation was short (3–8 h), there was no observable depletion of parent chemical and no production of any metabolites (as analysed by UV-HPLC or LC-MS). In order to detect metabolism in cellular test systems, experiments were adapted to incorporate radiolabelled caffeine at concentrations markedly higher (at least 1 mM) than those present *in vivo* (50–100 μM) and with extended incubation durations of at least 24 h (Berthou et al., 1989).

An extensive body of data in humans and experimental animals is available in the scientific literature (as reviewed in CIR, 2019; Arnaud, 2010; Müller and Jacobson, 2011; Monteiro et al., 2016) showing that caffeine is rapidly and extensively metabolised *in vivo* in mammalian liver. N-demethylation is the primary biotransformation observed resulting in cleavage of each of the three methyl groups to form the following three primary metabolites (Fig. 3): 84% paraxanthine (1,7-dimethylxanthine), 12% theobromine (3,7-dimethylxanthine) and 4% theophylline (1,3-dimethylxanthine). These three dimethylxanthine metabolites are shown to undergo further N-demethylation to produce methylxanthine metabolites possessing a xanthine scaffold with a single N-methyl group. Other biotransformations for the dimethylxanthine metabolites include oxidation to the xanthine scaffold to form a uric acid scaffold which may undergo amide bond hydrolysis resulting in cleavage of the dihydroimidazolone ring and N-acetylation (Gracia-Lor et al., 2017).

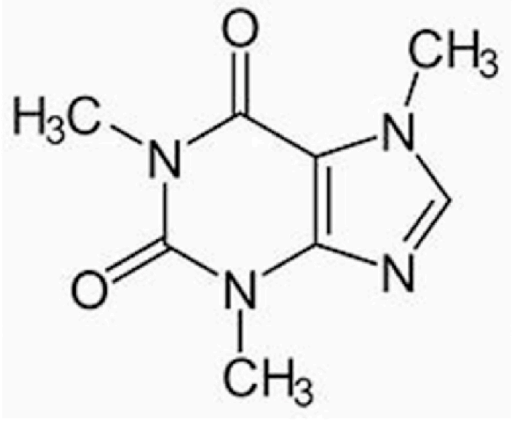


Fig. 2. Chemical structure of caffeine ($C_8H_{10}N_4O_2$; CAS 58-08-2; SMILES CN1C=NC2=C1C(=O)N(C(=O)N2C)C).

As already mentioned, it was assumed that no animal systemic toxicity data were available for caffeine.

Step 4: Analogue(s) a) identify, b) collate existing data, c) determine similarity hypothesis

3.1. Identify analogues

An initial screening of structurally similar analogues for caffeine, performed with the ChemTunes-ToxGPS software (https://www.mn-am.com/), resulted in the identification of 70 potential analogues that had a chemical structural similarity of above 70% based on the Tanimoto score, a similarity measure for comparing chemical structures. Structures are usually considered similar if the Tanimoto score is above 85%. Three chemicals had a score of greater than 90% similarity, i.e. theophylline, theobromine and paraxanthine (Table 2). But far more essential for the identification of analogues are other considerations such as presence of reactive groups, metabolism and physical chemical properties.

As mentioned in Step 2, caffeine consists of a xanthine scaffold with three N-methyl groups. Using the expert-judgement based framework for analogue selection by Wu et al. (2010), suitable analogues should possess the same molecular scaffold with the same functional groups. In this case, analogues should consist of a xanthine scaffold with at least one N-methyl substituent. As discussed above, metabolism studies demonstrate that N-demethylation is the primary metabolic transformation resulting in removal of each of the three methyl groups to produce three dimethylxanthine metabolites which may undergo further N-demethylation to form methylxanthine derivatives. Therefore, the three primary metabolites represent suitable analogues for caffeine since they are expected to display similar reactivity, metabolism and physical chemical properties to the target compound.

3.2. Collate existing data

A determinant of the outlined RAX approach was the availability of relevant data on physicochemical properties, biological activity and toxicity for the target chemical caffeine and the three highest scoring analogues (Tables 2–5). A literature and database search (OECD QSAR toolbox v4.3, scientific literature, COSMOS database) revealed predictive and *in vivo* systemic toxicity data in experimental animals for paraxanthine, theobromine and theophylline which supported the choice of the three chemical structure analogues.

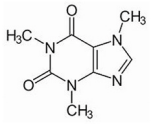
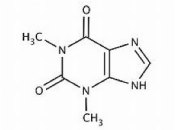
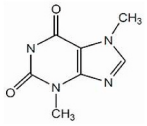
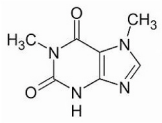
3.2.1. Physicochemical properties similarity

The selected analogues differ from caffeine by the deletion of a single N-methyl group resulting in largely similar physicochemical properties for all four compounds as shown in Table 2.

3.2.2. Metabolic similarity

As mentioned in Step 3, caffeine is rapidly and extensively metabolised *in vivo* in mammalian liver to three primary metabolites, i.e.

Table 2
Comparison of physicochemical and molecular properties of the target chemical caffeine and the analogues theophylline, theobromine and paraxanthine.

Name	Target chemical Caffeine	Analogue 1 Theophylline	Analogue 2 Theobromine	Analogue 3 Paraxanthine
CAS No.	58-08-2	58-55-9	83-67-0	611-59-6
Molecular formula	$C_8H_{10}N_4O_2$	$C_7H_8N_4O_2$	$C_7H_8N_4O_2$	$C_7H_8N_4O_2$
2D chemical structure				
Molecular Weight	194.19 g/mol	180.16 g/mol	180.16 g/mol	180.16 g/mol
Log Pow	−0.07	−0.02	−0.78	−0.63
Vapour Pressure (at 25°C)	0.00011999 Pa (non-volatile)	0.00006826 Pa (non-volatile)	0.00000015 Pa (non-volatile)	(non-volatile)
Water Solubility (at 25°C)	22000 mg/L	8300 mg/L	3300 mg/L	1000 mg/L
Lipinski rule	bioavailable	bioavailable	bioavailable	bioavailable
Living skin + RF absorption	Very high	Very high	Very high	Very high
Tanimoto coefficients	1.0 (target)	0.95	0.90	0.90

theophylline, theobromine and paraxanthine, which have also been identified as close structural analogues (Gracia-Lor et al., 2017).

Detailed information on the CYP-mediated metabolism of caffeine and theophylline was obtained using *in vitro* cell models expressing recombinant cytochrome P450 (CYP) enzymes (Ginsberg et al., 2004). The data demonstrate that the metabolism of caffeine by CYP is almost completely attributed to CYP1A2 (Table 3). The K_m and V_{max} kinetic parameters, together with the relative activity factor taking into account the relative abundance of each CYP in the liver, can be used to predict the formation of each metabolite. The models developed using these kinetic data provide an *in vitro* measurement of caffeine metabolism consistent with the results *in vivo* and represent an alternative assessment of caffeine metabolism further supporting the selection of the dixanthine metabolites as suitable analogues.

3.2.3. Biological similarity

Further scientific justification to increase the confidence in the analogue selection typically includes considerations of biological/mechanistic plausibility (Schultz et al., 2015). The U.S. Environmental Protection Agency's ToxCast program has screened thousands of chemicals for biological activity, primarily using high-throughput *in vitro* bioassays in order to differentiate pathway-specific from nonspecific effects (Fay et al., 2018). The target families with positive hits for caffeine included the cell cycle, nuclear receptor, DNA binding, GPCR (G-protein-coupled receptors) and esterase groups. All three analogues, theophylline, theobromine and paraxanthine, were found to have ToxCast *in vitro* assay data which confirmed the blockade of adenosine receptors by caffeine. However, the greatest mechanistic similarity was demonstrated between caffeine and theophylline which have shown activity in three of the assays that are related to adenosine binding such as nonselective binding to the adenosine A1 receptor, and selective binding to the adenosine A1 and the adenosine A2a receptor (Chavan et al., 2017).

3.2.4. Toxicological similarities

In silico prediction.

Using the OECD QSAR Toolbox v4.3, a search for predictive toxicology information was performed for the three analogues (theophylline, theobromine and paraxanthine) and the target chemical caffeine (Table 4). As it can be seen, the predictions were largely similar for caffeine and the three analogues with a few exceptions likely due to potential differences in reactivity. Overall, this further substantiated that a RAX approach was valid for these substances.

Legacy animal data on systemic toxicity for the analogues.

Literature search led to a number of repeated-dose and reproductive/developmental studies being described for the analogue substances (Table 5). Only studies that were scored as Klimisch 1 or 2 in terms of quality were considered valid to use (Klimisch et al., 1997).

Theophylline

Following repeated oral (diet) administration in rats over 14 weeks, theophylline caused nephropathy in male rats and a dose-dependent periarteritis in treated rats at all doses starting from 75 mg/kg bw/day; periarteritis was not observed in mice (NTP, 1998). Since the periarteritis is considered a rat-specific response to vasodilators (the pathogenesis of theophylline-induced vascular lesions may be a consequence of hemodynamic changes induced in the vascular wall, particular to the rat anatomy), it is of little, if any, relevance to humans (Nyska et al., 1998). Furthermore, this effect has not been associated with theophylline treatment in humans (OECD, 2001). In mice, adverse effects (mortality and reduced body weight) occurred at oral (gavage) doses of 150 mg/kg bw/day and above; therefore, a NOAEL for repeated-dose general toxicity after oral (gavage) exposure was set at 75 mg/kg bw/day.

Theophylline was not teratogenic in CD-1 rats at oral (diet) doses up to 259 mg/kg bw/day or in CD-1 mice at oral doses (drinking water) up to 396 mg/kg bw/day (OECD, 2001). At an oral dose of 218 (diet) and 396 mg/kg bw/day (drinking water), foetal toxicity was observed in rats and mice, respectively. However, foetal toxicity occurred only in the presence of maternal toxicity. Intravenous (IV) infusion of theophylline induced foetal toxicity in rabbits (foetal body weight reduction and increased incidence in cleft palate and 13th rib) at maternal toxic doses of 60 mg/kg bw/day (IV). However, this dose exceeded the effective therapeutic range of the substance (Shibata et al., 2000). Thus, for theophylline, the most sensitive NOAEL was defined at 30 mg/kg bw/day (IV) for both maternal and foetal toxicity. This pivotal study included measures of internal dose concentrations.

Theobromine

In a 13-week feeding study in rats, theobromine caused increased kidney weight and a reduced body weight gain, but only in male rats (Tarka et al., 1982). A NOAEL was set at 125 mg/kg bw/day. Reproductive/developmental studies with theobromine were performed in rats and rabbits (Theocorp Holding Company, 2010). Theobromine induced no foetal toxicity or teratogenicity in rats at an oral (diet) dose of 99 mg/kg bw/day, but there were signs of maternal toxicity (reduced food intake), with a NOAEL at 53 mg/kg bw/day. The oral NOAEL in rabbits for developmental toxicity was defined at 20–25 mg/kg bw (gavage and diet) on the basis of reduced foetal body weight and reduced food intake in dams. No foetal malformations were observed.

Paraxanthine

A single developmental toxicity study in mice was available for paraxanthine (York et al., 1986), which reported a dose-related increase in total malformations, primarily cleft palate and limb malformations, following intraperitoneal treatment at 175 and 300 mg/kg bw on

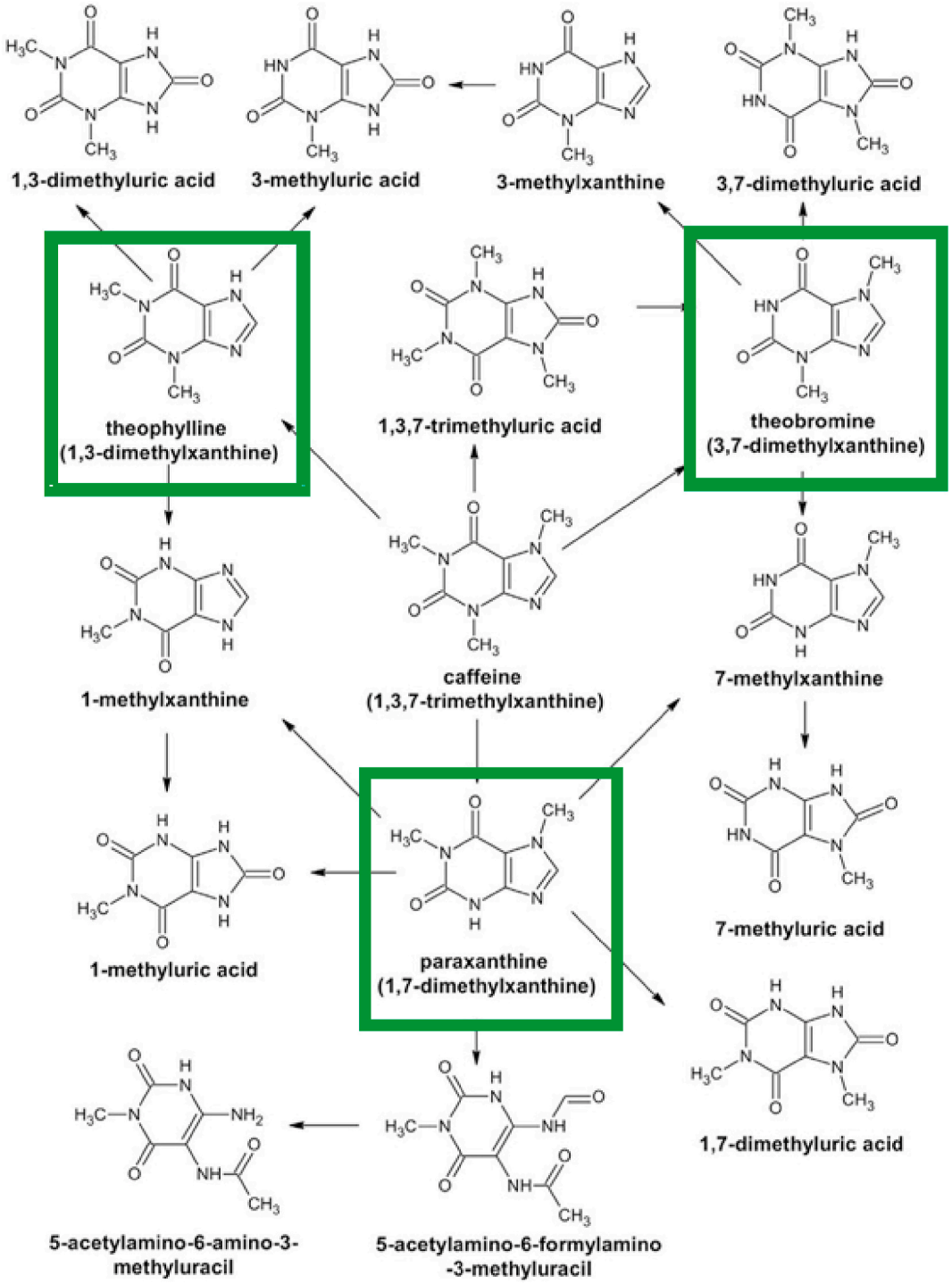


Fig. 3. Human urinary metabolites of caffeine (adapted from Gracia-Lor et al., 2017; major metabolites are highlighted).

gestational days 11 and 12. A NOAEL was not established.

3.3. Hypothesis for RAX

On the basis of the information collated so far, the hypothesis is that animal systemic toxicity data can be used from a structurally similar analogue of caffeine based upon a common metabolic pathway (Gracia-Lor et al., 2017; Schultz et al., 2015). Among the four possible RAX approaches, this scenario can be classified as using 'Chemical similarity involving metabolism (resulting in exposure to the same/similar substance(s))' (see Alexander-White et al., 2021). The use of a common metabolite approach is a cornerstone approach to RAX as per the ECHA Read-Across Assessment Framework (RAAF) (ECHA, 2017). In addition, this approach is most effective when the metabolites are major primary metabolites, as is the case here for caffeine. The hypothesis here is that all four chemicals, caffeine and the primary metabolites, have a common MOA.

Tier 0 exit: Step 4 → Step 8 Selection of a systemic toxicity point of departure (POD) for caffeine using RAX.

At the end of Tier 0, we considered whether we had enough information to move to step 8 and derive a POD. Several systemic toxicity animal data were gathered for the three analogues theophylline,

Table 3
Enzyme kinetics data from *in vitro* recombinant enzyme studies (CYP1A1, CYP1A2, CYP2E1 and CYP3A4) on the metabolic conversion of caffeine to paraxanthine, theobromine and theophylline (Ginsberg et al., 2004).

	Paraxanthine			Theobromine			Theophylline		
CYP Isoform	V_{max} [a]	K_m [b]	CL_{int} V_{max}/K_m	V_{max}	K_m	CL_{int} V_{max}/K_m	V_{max}	K_m	V_{max}/K_m
1A1	2.69	0.59	4.56	0.82	0.41	2.0	nd	nd	nd
1A2	30.5	0.19	161	3.0	0.16	18.8	1.12	0.25	4.5
2E1	Nd	nd	nd	0.48	1.44	0.33	0.36	0.84	0.43
3A4	Nd	nd	nd	nd	nd	nd	nd	nd	nd

[a] V_{max} units in moles of metabolite formed per h per mole CYP.
[b] K_m units in mmol/litre; nd = not detectable.

theobromine and paraxanthine, with more study data available for theophylline than for the other two analogues. The studies were judged to be of sufficient quality (scored as Klimisch 1 or 2; Klimisch et al., 1997).

Overall, the most sensitive toxic effect seen for theophylline, theobromine and paraxanthine was developmental toxicity such as reduction in foetal body weight gain and ossification as well as increased occurrence of supernumerary ribs. However, the foetal toxicity always occurred at maternally toxic doses and may therefore also be secondary to maternal toxicity (Khera, 1985).

The NOAEL of 30 mg/kg bw/day from the developmental toxicity study for theophylline in rabbits after intravenous infusion was selected as point of departure (POD) for the safety assessment. This was considered to be the most conservative NOAEL among the reported studies when taking into account exposure route and study duration. The adverse developmental effects occurred always in parallel with toxic effects in the dams such as reduced food intake and sometimes mortality. Overall, this choice of POD is considered to be conservative and consequently protective for human health, as the route of exposure is via IV infusion preventing oral or dermal metabolism, thus, resulting in 100% bioavailability by definition. In addition, IV administration generally produces high maximum plasma values thereby achieving toxic plasma levels which are not achieved following oral, and even more so, dermal administration of the same doses. In addition, the plasma levels needed to exert adverse developmental effects in humans are not attainable from ingesting large amounts of caffeine in foods and beverages. A profound review of 17 recent epidemiology (case-control and cohort) studies revealed no convincing evidence that moderate caffeine consumption by pregnant women ranging from 300 to 1000 mg per day throughout the entire pregnancy increases the risk of congenital malformations, miscarriage or birth defects (Brent et al., 2011; Wolde, 2014). According to EFSA (2015), prospective cohort studies show that caffeine intakes from all sources up to 200 mg per day consumed throughout the day by pregnant women in the general population do not give rise to safety concerns for the foetus.

The oral rabbit study data for theobromine indicates a NOAEL of about 25 mg/kg bw/day, similar to that for theophylline. The LOAEL were also similar (75 *vs* 60 mg/kg bw/day for theobromine and theophylline, respectively), but the theophylline data were preferred for the purposes of this illustrative case study, as they included internal dose metrics for theophylline.

One could in principle exit the framework at this point, since the selected POD appears suitably conservative. Using the external exposure dose metric, a risk assessment could be performed. However, at this point the margin of safety (MOS) = POD 30 mg/kg/day divided by the external exposure to cosmetics of 4.6 mg/kg/day (step 1 above) is 6.5 which is below the MOS of 100x considered acceptable here (SCCS, 2021). This does not mean caffeine is unsafe, it rather suggests that the data need to be further refined from a worst case scenario to a more realistic one, with an increasing level of confidence in both exposure and hazard assessment. For this purpose, the next step was to refine the exposure assessment by deriving an internal dose from the external exposure estimate using available *in vitro/in vivo* dermal absorption data (Tier 1).

4. Tier 1 – steps 5 and 6

Step 5: Systemic bioavailability/ADME of the target chemical and its analogues.

To refine the risk assessment for dermally applied cosmetic products, it is necessary to consider whether caffeine penetrates across the skin following dermal application and enters the systemic circulation.

The dermal penetration of caffeine has been studied extensively *in vitro* and *in vivo* (Table 6). The skin penetration values ranged from approximately 1 to 40% (1.25 ± 0.17% to 41 ± 20.03%) *in vitro* and from 2.5% to 62% (2.5–61.8 ± 5.4%) *in vivo*. The variable skin penetration rates depended on various factors, such as the applied vehicle/formulation, the test concentrations, site of the skin used in the tests, blocking of hair follicles, test duration and presence/absence of occlusion. These factors are considered equally important with respect to their impact on skin penetration. The objective was to take into account the situation under cosmetic use conditions and to be sufficiently conservative at the same time. Therefore, the highest *in vivo* skin penetration value in humans of 62% (Luo and Lane, 2015) was not regarded since that study used acetone as a vehicle, which is known to increase skin penetration and is not considered to be representative for a cosmetic formulation. Overall, a conservative value of 50% was selected for the safety assessment on the basis of the mean of the high-end *in vitro* (40%) and *in vivo* (60%) skin penetration data (Hewitt et al., 2020; Luo and Lane, 2015) (Table 6).

On the basis of an external cosmetics exposure estimate of 4.6 mg/kg bw/day (see Step 1) and a skin penetration rate of 50%, 2.3 mg/kg bw/day would stand for an assumed internal exposure metric. Taking into account the POD of 30 mg/kg/day, the risk assessment would result into a MOS of 30/2.3 = 13. Given that a default MOS of 100x is required to assure the safety of a substance (SCCS, 2021), further knowledge and refinement are needed, keeping in mind that this MOS is derived on the basis of worst-case assumptions concerning a potential human systemic exposure.

Step 6: Supporting a similar Mode of action (MOA) hypothesis

If it is possible to corroborate the similarity of the RAX analogues using a MOA hypothesis for the effects observed, this approach should be performed to refine the assumptions.

The pivotal MOA of methylxanthines is considered to involve a non-selective blocking of A1-and A2-adenosine receptors, thereby competitively inhibiting the action of adenosine in the cells. Also, adenosine receptor antagonism is considered the mode of action with most *in vivo* relevance with respect to the methylxanthine plasma concentrations reached through dietary intake. Since adenosine may exert multiple actions in the central nervous system, but also on the cardiovascular and other systems, this MOA may also be responsible for developmental

Table 4
QSAR Toolbox v 4.3. Predictive toxicity profilers for caffeine, theobromine, theophylline and paraxanthine.

	Analogue 1 Theophylline	Analogue 2 Theobromine	Analogue 3 Paraxanthine	Target chemical Caffeine
CAS number	58-55-9	83-67-0	611-59-6	58-08-2
SMILES	CN1C(=O)N(C)c2nc[nH]c2C1 = O	CN1C(=O)NC(=O)c2c1ncn2C	CN1C(=O)Nc2ncn(C)c2C1 = O	CN1C(=O)N(C)c2ncn(C)c2C1 = O
Profilers				
General Mechanistic				
DNA binding by OASIS	No alert found	No alert found	No alert found	No alert found
DNA binding by OECD	SN1 ≫ Iminium Ion Formation ≫ Aliphatic tertiary amines	No alert found	SN1 ≫ Iminium Ion Formation ≫ Aliphatic tertiary amines	SN1 ≫ Iminium Ion Formation ≫ Aliphatic tertiary amines
Toxic hazard classification by Cramer	High (Class III)	High (Class III)	High (Class III)	High (Class III)
Protein binding by OECD	Acylation ≫ Direct Acylation Involving a Leaving group ≫ Acetates	Acylation ≫ Direct Acylation Involving a Leaving group ≫ Acetates	Acylation ≫ Direct Acylation Involving a Leaving group ≫ Acetates	Acylation ≫ Direct Acylation Involving a Leaving group ≫ Acetates
Estrogen Receptor Binding	Non binder, without OH or NH2 group	Non binder, without OH or NH2 group	Non binder, without OH or NH2 group	Non binder, without OH or NH2 group
Protein binding by OASIS	No alert found	No alert found	No alert found	No alert found
Toxic hazard classification by Cramer (extended)	High (Class III)	High (Class III)	High (Class III)	High (Class III)
Endpoint Specific				
Skin irritation/ corrosion Exclusion rules by BfR	Group All Melting Point > 200 C; Group CN Melting Point > 180 C; Group CN Vapour Pressure < 0.001 Pa; Undefined	Group All Melting Point > 200 C; Group CN Melting Point > 180 C; Group CN Vapour Pressure < 0.001 Pa; Undefined	Group All Melting Point > 200 C; Group CN Melting Point > 180 C; Group CN Vapour Pressure < 0.001 Pa; Undefined	Group All Melting Point > 200 C; Group CN Melting Point > 180 C; Group CN Vapour Pressure < 0.001 Pa; Undefined
Oncologic Primary Classification	Not classified	Not classified	Not classified	Not classified
Acute aquatic toxicity classification by Verhaar (Modified)	Class 5 (Not possible to classify according to these rules)	Class 5 (Not possible to classify according to these rules)	Class 5 (Not possible to classify according to these rules)	Class 5 (Not possible to classify according to these rules)
Eye irritation/ corrosion Exclusion rules by BfR	Group All Melting Point > 200 C; Undefined	Group All Melting Point > 200 C; Undefined	Group All Melting Point > 200 C; Undefined	Group All Melting Point > 200 C; Undefined
DNA alerts for AMES by OASIS	No alert found	No alert found	No alert found	No alert found
Acute aquatic toxicity MOA by OASIS	Reactive unspecified	Reactive unspecified	Reactive unspecified	Reactive unspecified
rtER Expert System - USEPA	No alert found	No alert found	No alert found	No alert found
DNA alerts for CA and MNT by OASIS	No alert found	No alert found	No alert found	No alert found
Keratinocyte gene expression	High gene expression ≫ N-Acylamides	High gene expression ≫ N-Acylamides	High gene expression ≫ N-Acylamides	High gene expression ≫ N-Acylamides
DART scheme	Known precedent reproductive and developmental toxic potential; Purine and pyrimidine-like derivatives (7b)	Known precedent reproductive and developmental toxic potential; Purine and pyrimidine-like derivatives (7b)	Known precedent reproductive and developmental toxic potential; Purine and pyrimidine-like derivatives (7b)	Known precedent reproductive and developmental toxic potential; Purine and pyrimidine-like derivatives (7b)
Skin irritation/ corrosion Inclusion rules by BfR	Inclusion rules not met	Inclusion rules not met	Inclusion rules not met	Inclusion rules not met
Aquatic toxicity classification by ECOSAR	Carbonyl Ureas; Imidazoles	Carbonyl Ureas; Imidazoles	Carbonyl Ureas; Imidazoles	Carbonyl Ureas; Imidazoles
vitro mutagenicity (Ames test) alerts by ISS	No alert found	No alert found	No alert found	No alert found
Carcinogenicity (genotoxic and nongenotoxic) alerts by ISS	Structural alert for nongenotoxic carcinogenicity; Imidazole, benzimidazole (nongenotoxic)	No alert found	No alert found	No alert found
Respiratory sensitisation	No alert found	No alert found	No alert found	No alert found
Retinoic Acid Receptor Binding	Not possible to classify according to these rules	Not possible to classify according to these rules	Not possible to classify according to these rules	Not possible to classify according to these rules
Protein binding alerts for Chromosomal aberration by OASIS	AN2 ≫ Michael type addition to activated double bond of pyrimidine bases ≫ Pyrimidines and Purines; AN2 ≫ Shiff base formation with carbonyl group of pyrimidine or purine bases ≫ Pyrimidines and Purines	No alert found	No alert found	AN2 ≫ Michael type addition to activated double bond of pyrimidine bases ≫ Pyrimidines and Purines; AN2 ≫ Schiff base formation with carbonyl group of pyrimidine or purine bases ≫ Pyrimidines and Purines

(*continued on next page*)

Table 4 (*continued*)

in vivo mutagenicity (Micronucleus) alerts by ISS	H-acceptor-path3-H-acceptor	H-acceptor-path3-H-acceptor	H-acceptor-path3-H-acceptor	H-acceptor-path3-H-acceptor
Eye irritation/ corrosion Inclusion rules by BfR	Inclusion rules not met	Inclusion rules not met	Inclusion rules not met	Inclusion rules not met

Table 5
Repeated-dose and reproductive/developmental toxicity data for theophylline, theobromine and paraxanthine (Klimisch score 1 or 2).

Species	Test article	Route of exposure	Dosage & Duration	Results	Reference
F344/N Rat (10/sex)	Theophylline	Oral (diet)	0, 1000, 2000 and 4000 ppm (m: ca. 75, 125, 250 mg/kg bw/day; f: ca. 75, 125, and 275 mg/kg bw/day), 14 weeks	**LOAEL 75 mg/kg/day** – minimal nephropathy (m: 10/10; control: 10/10), mesenteric and pancreatic periarteritis (f: 1/10)	NTP (1998)
F344/N Rat (10/sex)	Theophylline	Oral (gavage)	0, 37.5, 75, and 150 mg/kg bw/day, 14 weeks	**LOAEL 37.5 mg/kg/day** – mesenteric periarteritis (m: 1/10, control 1/10; f: 2/10, control 0/10)	NTP (1998)
B6C3F1 Mouse (10/sex)	Theophylline	Oral (diet)	0, 1000, 2000 and 4000 ppm (m: ca. 175, 400, 800 mg/kg bw/day; f: ca. 225, 425, and 850 mg/kg bw/day), 14 weeks	**LOAEL 175 mg/kg bw/day (m); 225 mg/kg bw/day (f)** – reduced body weight	NTP (1998)
B6C3F1 Mouse (10/sex)	Theophylline	Oral (gavage)	0, 75, 150, and 300 mg/kg bw/day, 14 weeks	**NOAEL 75 mg/kg bw/day (m); 150 mg/kg bw/day (f)** – mortality at higher doses	NTP (1998)
Sprague-Dawley Rat (10/sex)	Theobromine	Oral (diet)	0, 200, 1000, 2000 ppm (equivalent to 0, 25, 125, and 250 mg/kg bw/day), 13 weeks (non-GLP)	**NOAEL 125 mg/kg bw/day** – at higher doses increased kidney weights and reduced body weight (m)	Tarka (1982)
Sprague-Dawley CD Rat	Theophylline	Oral (diet)	0, 1500, 3000, 4000 ppm (0, 124, 218, 259 mg/kg bw/day), gestational days (GD) 6–15	**NOAEL maternal toxicity/foetal toxicity 124 mg/kg bw/day, NOAEL teratogenicity 259 mg/kg bw/day**	OECD (2001)
CD-1 Mouse	Theophylline	Oral (drinking water)	0, 750, 1500 or 2000 ppm (282, 372, 396 mg/kg bw/day), GD 6-15	**NOAEL maternal toxicity/foetal toxicity 282 mg/kg bw/day, NOAEL teratogenicity 396 mg/kg bw/day**	OECD (2001)
Kbl: JW Rabbit (20 f/group)	Theophylline	Intravenous (IV, automatic infusion pump)	0, 15, 30 and 60 mg/kg bw/day (maternal plasma concentration C_P: 30, 56 and 106 μg/mL), GD 6-18	**NOAEL maternal toxicity 30 mg/kg bw/day, NOAEL foetal toxicity/teratogenicity 30 mg/kg bw/day**	Shibata et al. (2000)
Sprague-Dawley Rat	Theobromine	Oral (diet)	0, 625 and 1350 ppm (53 and 99 mg/kg bw/day), GD 6-19	**NOAEL maternal toxicity 53 mg/kg bw/day** – reduced food intake at the higher dose, **NOAEL foetal toxicity/teratogenicity 99 mg/kg bw/day**	Theocorp Holding Company, 2010
New Zealand Rabbit	Theobromine	Oral (gavage)	0, 25, 75, 125, 200 mg/kg bw/day, GD 6-29	**NOAEL maternal toxicity 75 mg/kg bw/day** - reduced food intake at higher doses, **NOAEL foetal toxicity/teratogenicity 25 mg/kg bw/day** – reduced foetal bw at higher doses	Theocorp Holding Company, 2010
New Zealand Rabbit	Theobromine	Oral (diet)	0, 625, 1250 and 1880 ppm (approx 0, 21, 41, or 63 mg/kg bw/day), GD 6-29	**NOAEL maternal toxicity 21 mg/kg bw/day** - reduced food intake at higher doses, **NOAEL foetal toxicity/teratogenicity 21 mg/kg bw/day**– reduced foetal bw at higher doses	Theocorp Holding Company, 2010
C57BL/6J Mouse (control and high dose: 16 f, low dose: 13 f)	Paraxanthine	Intra-peritoneal (IP)	0, 175 or 300 mg/kg bw (dissolved in deionized water), GD 11-12	**NOAEL maternal toxicity 175 mg/kg bw/day, NOAEL foetal toxicity/teratogenicity <175 mg/kg bw/day** - malformations	York et al. (1986)

toxicity, e.g. methylxanthines may affect neuronal growth and neuron interconnections during gestation and the neonatal period (Brent et al., 2011; Monteiro et al., 2016; Müller and Jacobson, 2011; Willson, 2018).

Given the close structural and metabolic similarity, it is hypothesised that these three analogue substances and the target chemical caffeine possess a similar MOA but with a different potency.

Inhibition of [^{3}H]-cyclohexyladenosine binding to rat cerebral cortical membranes *in vitro* by methylxanthines was investigated by Daly et al. (1983). Membranes were incubated with 1 nM [^{3}H] N_6-cyclohexyladenosine in the absence or presence of various methylxanthines. IC_{50} values were calculated as the means of three determinations. K_i values were estimated using the equation $K_i = IC_{50}/1 + [agonist]/EC_{50}$ (Daly et al., 1983) (Fig. 4). The substances behaved similarly in this assay and the specific relative potency factors (RPF) were calculated as follows: K_i Theophylline – Adenosine receptor 12 μM – RPF 1; K_i Paraxanthine – Adenosine receptor 30 μM – RPF 0.40 (1:2.5); Ki Caffeine – Adenosine receptor 50 μM – RPF 0.24 (1:4.16) and K_i Theobromine – Adenosine receptor 120 μM – RPF 0.10 (1:10).

These data suggest that theophylline is the most potent methylxanthine according to the inhibitory constant (K_i) for the A1-adenosine receptor in rat brain and therefore on the basis of this mechanistic evidence, theophylline is considered to be the most suitable compound to select as a 1:1 RAX analogue for caffeine. The mechanistic evidence here corroborates further the validity of the RAX approach and suggests that both caffeine and theophylline bind to the adenosine receptor at a comparable order of magnitude.

Tier 1 exit: Step 6 → Steps 8 – Performing a read-across (RAX) to derive a point of departure (POD) and Step 9 – Use of POD based on RAX to derive a margin of safety (MOS).

At the end of Tier 1, an assessment can be made as to whether the available data are sufficient to permit moving to Step 8. The MOA data in Step 6 suggest that the relative potency of caffeine to theophylline at

Table 6
Summary of Skin Penetration Studies with Caffeine (*in vitro* and *in vivo*).

Study Type	Caffeine concentration	Vehicle/Formulation	Fraction absorbed	Reference
In vitro (human skin)	1.08 μg/cm2	0.01M phosphate-buffered saline (PBS)	41 ± 20.03% (mass balance: 97.11 ± 2.21%)	Hewitt et al. (2020)
In vitro (human skin, 24 h)	1% (w/w), 260 mg/cm^2	W/O/W, O/W	1.25 ± 0.17% (W/O/W) 3.21 ± 0.18% (O/W)	Doucet et al. (1998)
In vitro (human skin, 24 h)	250 μg/cm^2	Ethanol (70%)	17% (unblocked hair follicles), 7% (blocked hair follicles)	Trauer et al. (2009)
In vitro (human skin, 72 h)	4 mg/mL, 10 μL/cm^2, repeated dosing (24 and 48 h)	ethanol:water (1:1, v/v)	Dermal delivery 13.69 ± 5.95%	Toner et al. (2009)
In vitro (human skin, 24 h)	1% (w/w), 10 μg/cm2	O/W, W/O	15–20%	Luo and Lane (2015)
In vivo Human subjects			2.5–3.3%	Zesch et al. (1979)
In vivo Human subjects (24 h), blocked and unblocked hair follicles	2.5%, 10 μg/cm^2	Ethanol and propylene glycol	Hair follicles unblocked: 24.9 ± 1.05% (receptor fluid) Hair follicles blocked: 12.4 ± 0.9% (receptor fluid)	Otberg et al. (2008) Trauer et al. (2009)
In vivo Human subjects (ventral forearm)	4 μg/cm^2	acetone	About 50%	Luo and Lane (2015)
In vivo Human subjects (forearm, non-occlusive patch, 24 h)	4 μg/2.5 cm^2	acetone	32.1 ± 4.2% (22–40 yr) 61.8 ± 5.4% (65–86 yr)	Luo and Lane (2015)

the adenosine receptor is 0.24. The MOS assessment is based on an intravenous POD (as 30 mg/kg bw/day) from the most suitable analogue theophylline, corrected for the potency difference to caffeine, and a worst-case internal exposure dose estimate of 2.3 mg/kg bw/day (see Step 5), resulting in (30/0.24)/2.3 = 54. Again, the MOS is lower than the desired 100fold. The MOS can be substantially increased by reducing the uncertainties associated with the exposure assessment. For this purpose, probabilistic modelling refinements, using consumer habits and practices data and real % use levels of caffeine in products could refine the external exposure assessment calculations from worst-case to more realistic values (Comiskey et al., 2015; Safford et al., 2015; Tozer et al., 2015, 2019).

In further steps (Tier 2), refinement in the risk assessment may involve targeted testing on the MOA hypothesis and PBK modelling to enhance human internal exposure estimates.

5. Tier 2 - steps 7 to 10

Step 7a: Targeted Testing using NAM assays – in this case exploring the possibility of endocrine-mediated activity involved in the observed developmental toxicity.

In Tier 1, the non-selective inhibition of the A1-adenosine receptor was considered to contribute to developmental toxicity effects secondary to maternal toxicity most probably caused by exaggerated pharmacological effects. However, further data are needed to exclude the possible involvement of endocrine-mediated mechanisms in this kind of toxicity.

Several *in silico* screening tools were used in order to look for alerts of potential endocrine activity related to estrogenic, androgenic, thyroidal and steroidogenic activities (EATS) activities (EFSA/ECHA Guidance, 2018). The results from the OECD QSAR toolbox v4.3, Annex 2 (estrogenic activity) showed that caffeine does not bind to the estrogen receptor. The Endocrine Disruptome tool (http://endocrinedisruptome.ki.si/) provides predictions of binding probabilities as a function of atomic-level information that is extracted from the three-dimensional structures of the ligand and the included nuclear receptors, and is a more insightful tool than other QSAR models that usually simply discriminate between binders and non-binders (Kolsek et al., 2014). This tool demonstrated for caffeine a 25- to 50% probability to act as an androgen receptor antagonist, and less than a 25% probability to bind to androgen, estrogen α and β, glucocorticoid, liver X α and β, PPAR α β ϒ, RXR α, thyroid α and β receptors. Another *in silico* tool, the VEGA platform, which provides tens of QSAR models to predict toxicity, ecotoxicity, environmental, and physicochemical properties of chemical

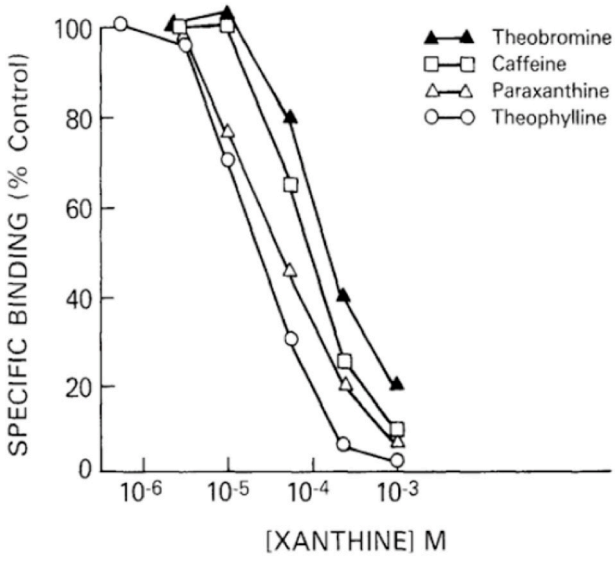


Fig. 4. Effect of theophylline, paraxanthine, caffeine and theobromine on A1-adenosine receptor systems in rat brain (extracted from Daly et al., 1983).

substances (https://www.vegahub.eu/download/), predicted that caffeine has no estrogenic activity.

Overall, the available *in silico* data for caffeine and the analogues theobromine, theophylline and paraxanthine, indicated only very low EATS receptor-mediated activities (probability of acting as NON-BINDER = >0.75), if any. Because of the absence of relevant *in silico* alerts, no hypothesis could be derived from these results. Therefore, additional generation of experimental *in vitro* data on potential EATS-related endocrine activity was not considered necessary. Indeed, some tests in the ToxCast database showed weak *in vitro* EATS-related endocrine activity for caffeine. However, these tests are screening assays, at the most OECD Conceptual Framework Level 2 studies, and not conducted according to GLP or OECD guidelines. Most tests revealed only borderline activity, where often only the highest concentration was above baseline, or it is stated that the result is potentially confounded by overfitting. In conclusion, these additional *in silico* data increased the confidence that there is no evidence of an EATS-related endocrine-mediated pathway contributing to developmental toxicity effects.

Step 7b: Kinetic refinements for target chemical and analogues

In Tier 1, the assumed systemic exposure dose of caffeine was calculated on the basis of a simple worst-case external exposure estimate and a dermal absorption rate derived from *in vitro* skin penetration data. To better estimate the internal concentration of caffeine after external aggregate exposure from the dermal and oral routes, physiologically based kinetic (PBK) modelling was performed.

Various guidance documents have been published for the application, use, best practice and reporting of PBK models (aka physiologically-based pharmacokinetic, PBPK) (WHO, 2010; USEPA, 2006; EMA, 2018). Additionally, a recent guidance was published by the OECD to address the credibility and increase confidence in these next generation PBK models for their intended purposes, in order to promote their acceptance and use in a regulatory context (Sachana, 2019; OECD, 2021). A number of recent reviews of good practice PBK modelling in environmental risk assessment are available (Clewell and Clewell, 2008; Campbell et al., 2012; Paini et al., 2019; Pletz et al., 2020). These case studies may help in establishing the best modelling strategy, as demonstrated by Gajewska et al. (2015) for caffeine and by Moxon et al. (2020) for coumarin exposure.

5.1. The PBK model structure and parameters

A human PBK model for caffeine was developed in Berkeley Madonna software (version 8.3.18; University of California, Berkeley, CA; www.berkeleymadonna.com). It is structured similarly to that reported by Gajewska et al. (2015), with perfusion-limited compartments for skin, liver, fat, lung, kidney, blood, and combined compartments for the remaining richly and slowly perfused tissues. The major difference with the Gajewska model was that a more simplified model was developed (i.e., single gastrointestinal and skin compartments), and the mass balance equations were corrected or rewritten. Since the results from this simplified model were very similar to the more complicated, multi-compartment model reported by Gajewska, no additional complexity was added. Consistent with common practice, tissue:venous equilibration was assumed, and the tissues were assumed to be well-mixed reservoirs. Exposure was characterised in exposed skin (dermal) and gastrointestinal (GI, oral) compartments, and metabolism was described as a first-order clearance process in the liver. The model structure is shown schematically in Fig. 5.

The physiological parameters used are shown in Table 7, and the chemical-specific parameters are shown in Table 8.

Tissue blood flows and volumes were set to values from Brown et al. (1997), except for skin. Skin volume was calculated as the product of surface area and average skin thickness. Skin thickness was taken from Brown et al. (1997) and skin surface area was calculated using the following allometric relationship from Livingston and Lee (2001):

$$SA = 0.1173 * BW^{0.6466}\ (m^2)$$

Blood flow to the exposed skin was calculated as the total skin blood flow adjusted by the ratio of volume exposed skin to volume of total skin. Exposed skin volume and blood flow was subtracted from the total skin to derive the parameters for the unexposed skin compartment.

Tissue:blood partition coefficients (PC) were estimated by Gajewska et al. (2015) using the algorithm of Schmitt (2008), except for skin. Skin:

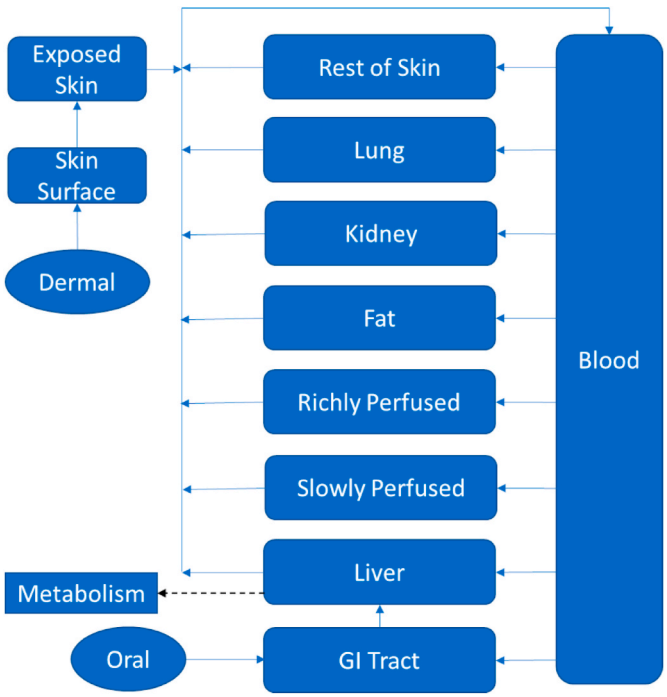


Fig. 5. PBK model schematic for caffeine showing the representation of the main organs considered with various sub-compartments in the skin and GI tract for oral and dermal exposure.

Table 7
Human PBK physiological parameter values for caffeine (source for values is Brown et al., 1997, unless otherwise specified).

Parameter	Units	Symbol	Value
Body weight[d]	kg	BW	–
Skin thickness	cm	Depth	0.1
Blood Flows (Fraction of Cardiac Output)[a]			
Cardiac Output	$L/h/kg^{0.75}$	QCc	15
Fat	1	QFc	0.052
Kidney	1	QKc	0.75
Liver	1	QLc	0.227
Skin	1	QSkc	0.058
Lung	1	QLuc	0.025
Volumes (fraction of BW)[b]			
Fat	1	VFc	0.21
Kidney	1	VKc	0.004
Liver	1	VLc	0.026
Lung	1	VLuc	0.008
Blood[c]	1	VAc, VVc	0.079
Hepatocellularity[e]	millions of hepatocytes per gram liver	hpgl	99

[a] Richly perfused blood flow = 70% of QC minus liver, kidney, and lung volumes; Slowly perfused blood flow = 30% of QC minus fat, and skin volumes.
[b] Richly perfused tissue volume = 70% of BW minus liver, kidney, and lung volumes. Slowly perfused tissue volume = 83.6% of BW minus fat, blood, and skin volumes.
[c] Blood is divided into 3/4 arterial and 1/4 venous.
[d] Simulation specific.
[e] Barter et al. (2007).

Table 8
Chemical-specific PBK modelling parameters for caffeine.

Parameter	Units	Symbol	Value	Source
Molecular weight Caffeine	g/mol	MWC	194.2	PubChem
Partition Coefficients				
Fat	1	PF	0.68	Schmitt (2008)
Kidney	1	PK	3.76	Schmitt (2008)
Liver	1	PL	4.25	Schmitt (2008)
Lung	1	PLu	1.23	Schmitt (2008)
Rich	1	PR	2.4	Schmitt (2008)
Slow	1	PS	0.995	Schmitt (2008)
Blood and Plasma				
Fraction unbound in blood	%	fub	96	Lave et al. (1997)
Fraction unbound in plasma	%	fup	68	Lelo et al. (1986)
Blood:plasma ratio	%	RBP	71	fup/fub
Oral absorption				
GI - > liver	1/h	Ka	1.6	fit
Dermal absorption				
Fraction available	%	FracAvail	50	Géniès et al. (2019), Hewitt et al. (2020)
Permeability	cm/h	Kp	4.10×10^{-4}	Doucet et al. (1998), fit
Metabolism				
Hepatocyte clearance	uL/min/million cells	hep_Cl_{int}_invitro	0.68	fit

blood PC was estimated as a weighted average of liver and fat PCs (i.e. 0.7*PL + 0.3*PF).

The caffeine PBK model simulates two routes of exposure, oral and dermal.

1. Caffeine exposure from other sources includes oral intake from coffee, tea, energy drinks, cola and chocolate (EFSA, 2015). The worst-case (maximum 95th percentile) of caffeine intake from all food/drink sources is estimated to be 648 mg/person/day (10.8 mg/kg bw/day) for adults (18 to <65 yr) and 786 mg/person/day (13.1 mg/kg bw/day) for elderly (≥65 yr). The latter value was used as input into PBK modelling for the oral route.

Oral exposure is modelled using a 1-compartment, first-order absorption model, with the GI tract acting as a reservoir for the oral dose. All oral doses are simulated as bolus doses (i.e. total substance ingested at once per dosing event). The blood flow from the GI enters the liver via the portal vein.

2. Dermal exposure is modelled using a skin surface compartment to house the applied dose in terms of the volume and surface area, and a single skin compartment. Transfer to the systemic circulation occurs in the skin compartment. The skin is separated into exposed and unexposed compartments. Dermal absorption is driven by a permeability coefficient for uptake from the surface into the skin (units of cm/hr), the exposure area and volume of application, the amount of chemical applied, duration of application, and the fraction absorbed. Transfer from the skin to the blood is modelled assuming a well-mixed, blood flow-limited exposed skin compartment and a skin:blood partition coefficient.

Measured data were required to first build a PBK model, which was parameterised and then further data were used to simulate and test the model. To simulate the experimental dermal exposure data from Otberg et al. (2008), applied volume and concentration was calculated from details reported in the exposure methodology. For the consumer whole-body cosmetic exposure scenario, the applied volume was assumed to be 4 mL (Troutman et al., 2015), the surface area exposed was taken to be the whole skin surface area, and the concentration was assumed to be 2% in lotion. The fraction available for dermal absorption was extrapolated based on a 2% concentration and 50% dermal penetration (Hewitt et al., 2020; Luo and Lane, 2015). Specifically, assuming a lotion density of 1 g/mL, 2% corresponded to 20 mg/mL, which was higher than the greatest test concentration of 10 mg/mL. Inspection of the graph showed decreasing penetration into the medium as concentration increased. The 10 mg/mL absorption, which was close to 50%, was used as a conservative value to estimate availability at the estimated simulated concentration of 20 mg/mL.

Some of the parameters used in Tier 0 to review similarity were also used as input parameters in PBK modelling. Also, a permeability coefficient (Kp) was needed in the PBK model to calculate the rate of absorption. Doucet et al. (1998) reported a Kp of 6.0×10^{-4} cm/h in frozen female abdominal excised skin in a simple oil/water emulsion. This value was close to the value of 2.1×10^{-4} cm/h measured by Dias et al. (1999) using a saturated solution of caffeine in water. No human *in vitro* or *in vivo* dermal delivery data for theobromine or paraxanthine were found. For theophylline an *in vitro* study in human skin reported a Kp value of 2.1×10^{-5} (Kopecna et al., 2017). This value was later fitted to the data.

The intrinsic hepatic clearance and the rate of dermal penetration and oral absorption were fitted to experimental data collected in controlled oral and dermal human exposures (Denaro et al., 1991; Lelo et al., 1986; Otberg et al., 2008) (Figs. 6–8). The parameters were fit simultaneously to the individual data sets, and the average fitted value was used for all subsequent simulations. The skin permeability measured by Doucet et al. (1998) was used as the starting point for the dermal uptake, and adjusted upward to 6.0×10^{-3} to obtain a fit to the human volunteer data. Doucet et al. (1998) used frozen female abdominal skin tissue, while Otberg et al. (2008) exposed male chests with "pronounced terminal hair on the chest". The adjustment to the *in vitro* value was considered reasonable as there are known absorption differences depending on vehicle and extent of hair follicle density across body locations.

The liver clearance was calculated by scaling the *in vitro* intrinsic clearance value to the whole body. The liver clearance rate was calculated as the μL/min/million cells, times hepatocellularity, times the volume of the liver:

$$CLint(L/h) = CLint\_invitro * hpgl * 10^{-6}(L/\mu L) * 60\ (\min/h) * 10^{3}(g/kg) * VLc * BW(kg)$$

where CL_{int}*in vitro* was the μL/min/million cells cleared *in vitro*, hpgl was the number of million hepatocytes per gram of liver, and VLc*BW the volume of the liver.

A local sensitivity analysis for model parameters was conducted using the built-in tool in the Berkeley Madonna software. The sensitivity coefficients were then normalised by the output and input parameter values according to the following equation:

$$Normalised\ Sensitivity\ Coefficient = \frac{\Delta Y/Y}{\Delta X/X}$$

where Y was the output (i.e., C_{max} or AUC), X was the input parameter (e.g., Ka, Kp), ΔX the change in the parameter value, and ΔY the resulting change in the output value. Normalization of the sensitivity coefficients was necessary to make comparisons across parameters of different scales (Clewell et al., 1994).

5.2. Estimation of internal exposure using PBK modelling

A qualitative evaluation of the agreement between experimental plasma concentration and simulations was conducted through visual inspection of the time-course concentration curves (USEPA, 2006). Good

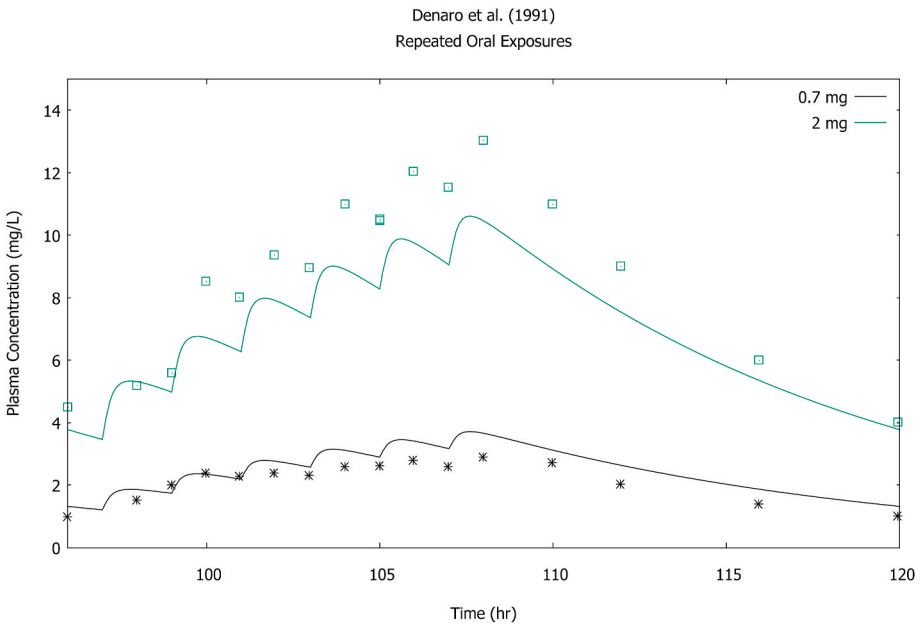


Fig. 6. Caffeine PBK simulations compared to measured repeated oral exposure data in human volunteers (Denaro et al., 1991). Caffeine plasma concentrations (black stars, green squares) were determined following ingestion of 6 cups of caffeinated coffee (0.7 and 2 mg/kg caffeine) over 5 days. Since the body weight (BW) was not reported, a default value of 70 kg was used. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

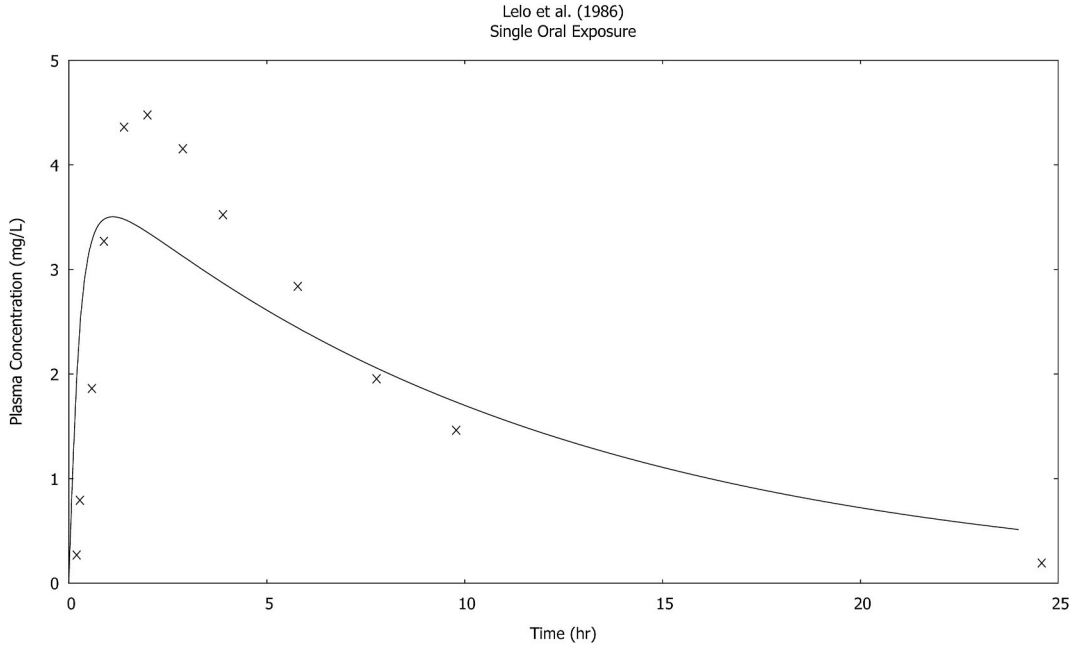


Fig. 7. Caffeine PBK simulations compared to measured oral exposure data in human volunteers (Lelo et al., 1986). Caffeine plasma concentrations (black crosses) were determined following ingestion of 3.25 mg/kg bw caffeine in a gelatin capsule. BW = 83 kg.

agreement was obtained for both the oral and dermal route, with model predictions generally within a factor of two of the data (WHO 2010) (Figs. 6–9). Only the hepatocyte intrinsic clearance (CLint), and oral (Ka) and dermal (Kp) absorption coefficients were fitted to experimental data, while all remaining parameters were obtained from the literature or data collected by Cosmetics Europe (Géniès et al., 2019; Hewitt et al., 2020). The values of CL_{int} obtained from fitting to the repeated oral (Denaro et al., 1991) and dermal data (Otberg et al., 2008) in human volunteers were very similar (0.7 vs 0.5 μl/min/million hepatocytes). A value of 1.4 μl/min/million hepatocytes was reported by Lelo et al. (1986), which is also close to the fitted values.

A simulation of the worst-case exposure estimated for caffeine via oral and dermal routes was conducted (Fig. 9). Upper bound oral (13.1 mg/kg/d) and dermal (4.6 mg/kg/d) exposure estimates (see Tier 0) were determined and used as input to the model. Twice daily exposures were simulated, 12 h apart, as bolus ingestions or dermal applications. A 4 mL volume was used to simulate whole body exposure to lotion, as reported by Troutman et al. (2015). Caffeine concentration in the dermal formulation was assumed to be 2%, which corresponds to approximately 20 mg/mL caffeine. Using the available dermal penetration data, an average value of 50% dermal absorption was assumed for caffeine (see Tier 1, Step 5). A default body weight of 60 kg was used as recommended in the SCCS Notes of Guidance (SCCS, 2021).

The simulated internal dose metrics (Cmax and AUC) are shown in Table 9. The dosing simulation was run for 3 (simulated) days to achieve steady periodicity, as shown in Fig. 9. The maximum concentration in

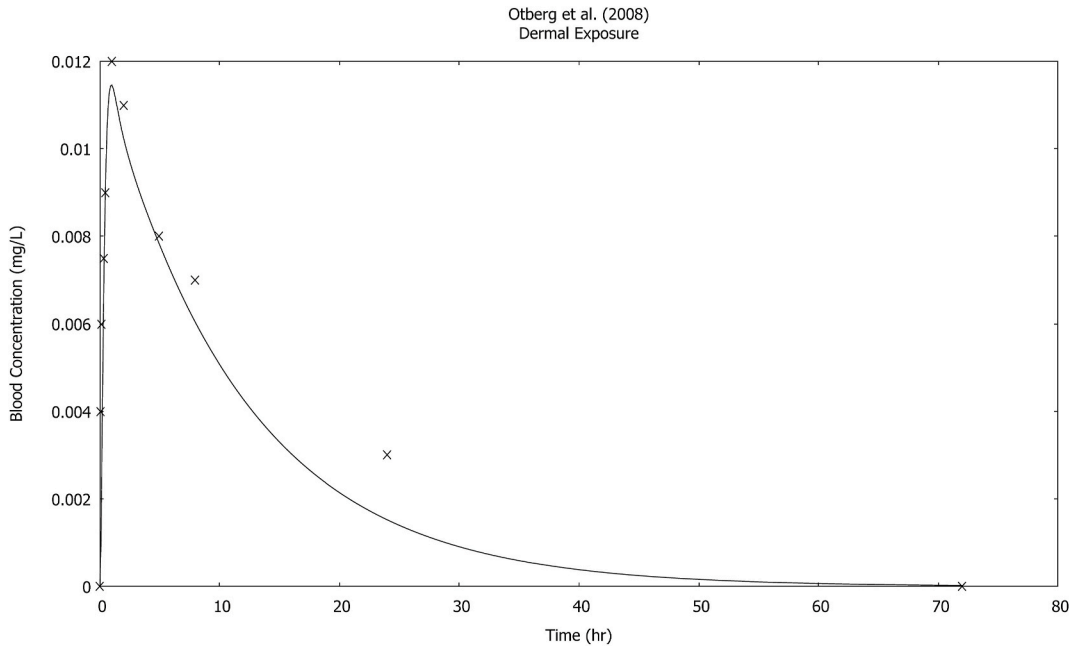


Fig. 8. Caffeine PBK simulations compared to measured dermal exposure data in human volunteers (Otberg et al., 2008). Caffeine plasma concentrations (black crosses) were determined following dermal application of 50 mg (0.05 mL) of an ethanol/propylene glycol formulation containing 2.5% caffeine to 25 cm^2 of the chest of 6 male volunteers (1.25 mg caffeine). Since the body weight (BW) was not reported, a default value of 70 kg was used.

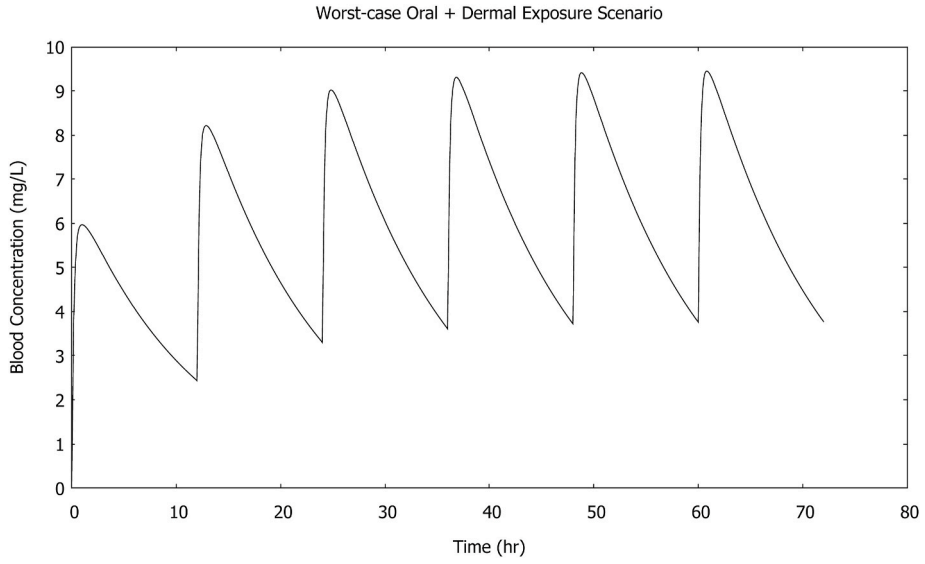


Fig. 9. PBK simulation of internal plasma concentration of caffeine following estimated oral and whole body dermal exposure. Oral exposure = 13.1 mg/kg/d, whole body dermal exposure = 4.6 mg/kg/d in 4 mL of product, twice daily exposure (2 × 2.3 mg/kg/d). BW = 60 kg. Cmax = 9.4 mg/L, AUC = 150 mg*h/L, Cavg = AUC/24 = 6.4 mg/L.

blood (Cmax) was 9.4 mg/L, the daily area under the concentration curve (AUC, calculated over the time period from 48 h to 72 h) was 150 mg*h/L, and the average daily concentration (Cavg) was 6.4 mg/L. From sensitivity analyses (data not shown), the model is generally positively affected (increased Cmax) by increases in parameters driving absorption (FracAvail and Ka), and negatively (decreased Cmax) by increases in parameters driving clearance (e.g., liver parameters and intrinsic clearance).

Table 9 shows the calculation of Cmax and AUC from the PBK model. The fraction available for dermal absorption was set to 50% based on extrapolation of the trend found using *in vitro* testing with human skin. In humans, when using acetone as a vehicle the fraction absorbed increased to 62%, but this was not considered to be representative for a cosmetic formulation. However, in order to examine the sensitivity of the model, it was run again with a skin penetration rate of 62%. The impact of this increase in skin penetration was low with an approximate 6% increase in both C_{max} and AUC.

At the end of Tier 2 with new PBK data, we can reiterate Step 8 (as in Tier 0 and Tier 1) and use the same POD based on RAX from theophylline to derive a margin of safety (MOS) in Step 9.

Step 9: Performing a margin of safety (MOS) evaluation

Usually, we can calculate a margin of safety (MOS) through dividing

Table 9
Summary of the worst-case cosmetic exposure scenario. Daily oral and dermal doses are split into 2 equal doses, administered at 12-h intervals.

Exposure	50% Skin penetration	62% Skin penetration
Single Oral Dose (mg/kg)	6.55	6.55
Single Dermal Dose (mg/kg/d)	2.3	2.3
Dermal Exposure Area cm^2	16560	16560
Doses/day	2	2
Daily Oral Dose (mg/kg/d)	13.1	13.1
Daily Dermal Dose (mg/d)	4.6	4.6
Internal Dose Metrics		
Cmax (mg/L)	9.4	10
AUC (mg-h/L)	150	160
Cavg (mg/L)	6.4	6.8

the POD by the exposure estimate for the target chemical.

In this case study, a margin of internal exposure (MOIE) has been defined for caffeine using a PBK model which estimated blood concentrations derived from external caffeine exposure data in humans. A MOIE differs from a traditional margin of safety (MOS) in that it is calculated as the ratio of a measure of internal exposure, such as blood/plasma concentration or target-tissue dose, rather than a measure of external exposure concentration, total bolus dose or ingested dose (Bessems et al., 2017).

The ability to rely on a kinetic measure of internal rather than external exposure reduces the uncertainty in the risk assessment by incorporating chemical-specific information on the uptake, distribution, metabolism and excretion in both experimental animal and human as defined using species-specific physiological parameters (Clewell et al., 2008). In particular, calculation of internal exposure with a PBK model can be used to replace the default uncertainty factor of 4 for interspecies differences in toxicokinetics (WHO, 2010). The USEPA follows this practice in determining Reference Concentrations and Reference Doses (USEPA, 1994 and 2011). Thus, a MOIE of 25 is considered equivalent to the default MOS of 100, but with greater precision for the target chemical.

On the basis of experimental data, the internal exposure (plasma concentration, C_p) at the most conservative POD from animal studies with theophylline, the closest analogue to caffeine (NOAEL 30 mg/kg bw/d, IV injection), was 56 μg/mL, equivalent to 311 μM (Shibata et al., 2000). According to PBK modelling, the worst-case internal caffeine exposure following human aggregate dermal (cosmetic products) and oral (food/drink) caffeine exposure (C_{max}) was estimated as 9.4 mg/L, equivalent to 48 μM. Using the MOA data related to the adenosine receptor, the relative potency of caffeine is lower than that of theophylline and a relative potency factor of 0.24 can be applied (see Step 6). Thus, the MOIE is calculated as follows:

$$MOIE = C_p \text{ NOAEL animal study} / (C_{max} \text{ human x RPF}) = 311\ \mu M / 48\ \mu M \times 0.24 = 27$$

In addition, a comparison of suitable *in vivo* caffeine animal data with the data of its most potent analogue theophylline supported the reliability of the RAX part of the outlined NGRA. When comparing the findings from oral 3-month studies (same rat strain, no bolus exposure), caffeine appeared to be slightly less toxic compared to theophylline, its most potent analogue, i.e., caffeine NOAEL of 151 mg/kg bw (drinking water) *vs* theophylline LOAEL of 75 mg/kg bw (diet) (Tables 5 and 10). The NOAELs from the developmental toxicity studies were derived from different species and exposure routes (theophylline, rabbit NOAEL 30 mg/kg bw IV *vs* caffeine, rat NOAEL 40 mg/kg bw oral gavage), but appear to indicate a similar dose range of toxicity. However, the calculation of a MoIE on the basis of safety data with caffeine, i.e. direct comparison of the estimated internal human exposure following dermal cosmetic use and the POD/NOAEL from a caffeine animal toxicity study, is not possible since in none of the available animal studies internal compound exposure was assessed. Due to insufficient data and resources, the generation of an animal PBK model for caffeine was not possible. Therefore, the read-across approach with the theophylline data was so valuable.

Step 10 Assessing the level of confidence in the safety assessment approach

Overall, the level of confidence in the key aspects of the safety assessment was considered medium to high (Table 11). On the basis of the common MOA between caffeine and the closest structural analogue theophylline, internal plasma levels derived from the most conservative NOAEL among the available animal systemic toxicity studies with theophylline were used as POD. The final MOIE calculation was considered sufficiently conservative since it was based on estimated internal exposure with a PBK model, rather than on external doses with inherent uncertainty related to route-to-route extrapolation and species/strain differences. Thus, calculation of internal exposures with a PBK model can be used to replace the default uncertainty factor of 4 for interspecies differences in toxicokinetics, as the data are based on physiological parameters and exposure data in humans.

Consequently, the resulting MOS/MOIE of 27, with all of the in-built conservatism in this case study, demonstrates the safety of caffeine from combined exposure to cosmetics and food/drinks for the consumer.

Table 10
Repeated-dose and reproductive/developmental toxicity data for caffeine (Klimisch score 2).

Species	Route of exposure	Dosage & Duration	Results	Reference
F344/N Rat (12/sex)	Oral (drinking water)	0, 188, 375, 750, 1500, 3000 ppm (m: 19.7, 42, 85.4, 151, 272 mg/kg/day; f: 23, 51, 104, 174, 287 mg/kg bw/d), 90 days	NOAEL 151 mg/kg/day – significant body weight reduction at higher doses (20–26%), no other findings of toxicological relevance	OECD (2002)
Sprague-Dawley CD Rat (20/group)	Oral (gavage)	0, 40, and 80 mg/kg/day, gestation days (GD) 1–19	LOAEL maternal toxicity 40 mg/kg/day – reduced bw gain, NOAEL foetal toxicity 40 mg/kg/day – reduced foetal bw gain, NOAEL teratogenicity 80 mg/kg/day	OECD (2002)

6. Discussion

We want to emphasise that risk assessment is always based on models, either on animal models as in the traditional approach or on *in silico/in vitro* NAM as in the NGRA approach. And we should keep in mind: *All models are wrong, but some are useful* (Box, 1976).

In this case study on caffeine we set out to validate whether the application of the 10-step RAX framework for NGRA as described in Alexander-White et al. (2021) is a realistic approach to assure the safety of a substance in the absence of animal toxicity data.

Similar to the traditional safety assessment on cosmetic ingredients, the NGRA approach relies on two key parts, i.e. exposure and hazard assessment. For all safety assessments, it is essential to determine first as precisely as possible the external exposure to the respective cosmetic substance while also taking into account exposure from non-cosmetic uses, and then to estimate a maximum internal exposure. For caffeine, our first rough estimate of internal exposure was to take a conservative high-end skin penetration rate of 50% which, however, did not result

Table 11
Qualitative assessment of the level of confidence in the NGRA approach.

Factor	Level of Confidence (low, medium, high*)	Comment
Overall assessment of the hypothesis used for RAX, i.e. use of animal systemic toxicity data from a structurally similar caffeine analogue based upon a common metabolic pathway to support the risk assessment	+++	Three chemically similar analogues of caffeine (theophylline, theobromine, paraxanthine) were selected based on structure, physicochemical properties, ADME data and *in vitro* data determined using ToxCast assays. It is assumed that all four chemicals have a common mode of action (MOA).
Structural similarity based on Tanimoto score, molecular scaffold, metabolic transformation	+++ +++ +++	Caffeine and its three selected analogues reveal a high level of structural similarity based on • the Tanimoto score (>0.9) (Madden et al., 2020) • the same xanthine scaffold with at least one N-methyl substituent as functional group • N-demethylation as primary metabolic transformation
Similarity of physicochemical properties	+++ +++ ++ +++	Caffeine and its three selected analogues (theophylline, theobromine, paraxanthine) have sufficiently similar physicochemical properties, i.e. • similar MW and logP range • similar negligible volatility • sparingly to slightly water soluble • all are predicted to be bioavailable and have a good skin penetration
ADME similarity	+++ ++ ++	A direct comparison of the pharmacokinetics of caffeine, theophylline, theobromine and paraxanthine after oral intake (250–270 mg) in man demonstrated comparable plasma concentration-time curves (Lelo et al., 1986): • After oral intake, caffeine and its three selected analogues are almost completely absorbed (up to 99%) into the bloodstream • Metabolism mostly performed by phase I (cytochrome P450 CYP) enzymes, mainly CYP1A2 (almost 90% of caffeine metabolism), N-demethylation to paraxanthine (84%), theobromine (about 12%) and theophylline (about 4%) or oxidation to urate and/or hydration to diaminouracil metabolites • Similar plasma clearance between caffeine and paraxanthine (about 2 mL/kg/min) vs theobromine and theophylline (about 1 mL/kg/min)
Mode of action (MOA)	+++	The common pivotal MOA of methylxanthines is considered to involve a non-selective blocking of A1- and A2-adenosine receptors, thereby competitively inhibiting the action of adenosine in the cells. Since adenosine may exert multiple actions in the CNS, but also on the cardiovascular and other systems, this MOA may also be responsible for developmental toxicity, e.g. methylxanthines may affect neuronal growth and neuron interconnections during gestation and the neonatal period.
Similarity of other supportive data such as ToxCast	++	The enzymatic and receptor signaling assays from the US EPA ToxCast database identified adequate biological similarities of caffeine with the selected analogues, with theophylline having the greatest mechanistic similarity in terms of interaction with adenosine receptors.
Number of analogues used for the read across	++	The number of analogues selected for the read-across was considered sufficient.
Quality of the toxicity endpoint data used for the read across	+++	Animal studies of acceptable quality (Klimisch score 1 and 2) were available for the selected analogues.
Similarity of the toxicity endpoint data (among source chemicals)	++	The overall effect profile was comparable, but with differences in potency
External exposure assessment	++	The external exposure assessment for caffeine was a worst-case estimate of aggregated dermal and oral exposure. Therefore, confidence is only medium since the measure of external exposure is considered to largely overestimate the external exposure under real consumer conditions
Application of a PBK model to derive an internal exposure estimate for the risk assessment	++ +++	The PBK model simulations agreed reasonably well across multiple oral and dermal exposure data sets using a common set of parameter values. It tended to underpredict the plasma concentration at the higher oral doses, though it was consistently within a factor of 2 from the data. This is likely due to saturation of oxidative metabolism, and there are also potential vehicle effects that are not being captured by the model. A more complex description of the oral absorption model could potentially improve the accuracy of the simulations at high doses. The estimated internal exposure based on the PBK model substantially increased the confidence in the final MOS/MOIE calculation due to avoidance of external doses with inherent uncertainty related to route-to-route extrapolation and species/strain differences

+, ++, +++ = parameter likely to cause low, medium, high confidence.

into an acceptable MOS. A subsequent refinement of the potential human internal exposure was achieved by PBK modelling on the basis of actual human data after oral, as well as dermal exposure, assuming external worst-case aggregate exposure to caffeine-containing products by both the oral (food/drinks) and dermal (cosmetics) routes.

With respect to the hazard assessment part, RAX is considered a cornerstone of the NGRA framework in order to derive a POD for the safety assessment. Therefore, potential caffeine analogues were identified by initial screening on the basis of structural similarity with the help of specific computer softwares. However, suitable analogues should possess the same molecular scaffold with the same functional groups (Wu et al., 2010), in this case a xanthine scaffold with at least one N-methyl substituent. The final choice of analogues was strengthened by data demonstrating a satisfactory degree of similarity with respect to physicochemical properties, metabolism, enzymatic and receptor signaling assays as well as alerts for toxicological endpoints.

As a result of the available NAM data (*in silico* and *in vitro*), we regarded the caffeine analogues theophylline, theobromine and paraxanthine forming a robust analogue category for the application of the RAX approach for the following reasons:

a) very high degree of structural similarity, i.e. based on the Tanimoto score (≥0.9) as well as on the same xanthine scaffold with at least one N-methyl substituent as functional group;
b) sufficiently similar physicochemical properties;
c) caffeine is metabolised *in vivo* to those three analogues with paraxanthine being the major metabolite;
d) the enzymatic and receptor signaling assays from the US EPA ToxCast database identified adequate biological similarities of caffeine with its analogues, with theophylline having the greatest mechanistic similarity;
e) an *in silico* safety alert tool, the OECD QSAR toolbox v4.3, predicted largely similar toxicological characteristics for caffeine and those three analogues;

Systemic toxicity animal data of sufficient quality were gathered for two caffeine analogues, i.e. theophylline and theobromine. Overall, the most sensitive toxic effect seen for theophylline and theobromine was foetal toxicity which may also be related to the maternal toxicity occurring at the same doses (Khera, 1985). The selected POD of 30 mg/kg bw/day was derived from a developmental toxicity study with theophylline. This POD was considered to be highly conservative and consequently protective for human health, as the route of exposure was via IV infusion preventing oral or dermal metabolism and resulting in 100% bioavailability by definition.

Given their close structural and metabolic similarity, our hypothesis was that caffeine and its analogues may have a common MOA but with a different potency. The pivotal common MOA of caffeine and its analogues was considered to be adenosine receptor antagonism with theophylline having the highest relative potency. This MOA may also contribute to the observed maternal toxicity (Monteiro et al., 2019).

In order to strengthen our hypothesis, further NAM data were generated with the help of several *in silico* screening tools which increased the confidence that there is no evidence of an endocrine-mediated pathway underlying the developmental toxicity effects (Schuhmacher-Wolz et al., 2017).

Cosmetics are not drugs. Therefore, for many cosmetic ingredients or analogues of cosmetic ingredients it will be difficult or even impossible to define a MOA, given that cosmetics rarely contain active ingredients at pharmacologically relevant concentrations. Then, showing qualitative and quantitative concordance of the *in vivo* and *in vitro* biological data among the analogues may be the only option to justify the analogue selection.

Here, we should also keep in mind that the traditional safety assessment of cosmetic substances on the basis of animal data includes multiple levels of uncertainty, such as.

- rodent toxicity data *vs* potential human toxicity
- rodent oral gavage toxicity data *vs* potential human toxicity after dermal exposure
- pharmacokinetic differences of dermal (human) *vs* rodent (oral) exposure data
- the uncertainty of applying an *in vitro* skin penetration model to estimate a hypothetical human systemic exposure

So, also the traditional risk assessment approach based on animal data poses models upon models. Thus, in order to take account of the inherent uncertainties of these models, safety assessors usually apply safety factors or generate additional data.

In our tiered exposure-driven and evidence-based framework, NAM were used to reduce the uncertainties of the RAX approach with respect to the differences in metabolism and kinetics between species, and between target and analogues substances. NAM were also used to strengthen the MOA hypothesis. And, as in the traditional risk assessment, appropriate safety factors were applied to account for remaining uncertainties.

A key asset for the NGRA on caffeine was the determination of the internal dose metric via PBK modelling. This model resulted in a more realistic estimation of a kinetic-based MOIE. The application of caffeine-specific kinetic information on the basis of human physiological parameters considerably reduced previous uncertainties concerning internal exposure estimates that were based on route-to-route extrapolations and interspecies kinetic differences. Once again, the traditional way to assess human safety after dermal exposure on the basis of oral animal toxicity data represents multiple models of uncertain values. Here, our approach suggests that a lower MOS may be acceptable simply by replacing the default uncertainty factor of 4 for interspecies kinetic differences (WHO, 2010).

Overall, our case study demonstrates the added value of NAM, such as *in silico* tools, *in vitro* ADME, PBK modelling and RAX. Whereas a single tool of this kit may give an alert, which does not necessarily mean the lack of safety of a chemical, our approach combined the power of the individual models. Our results demonstrated an outstanding degree of consistency and clearly identified potential hazards as well as the absence of risk to human health. In our view, these novel approaches open new perspectives on adapting and improving traditional safety assessment schemes. The applied RAX approach, i.e. using the theophylline POD for caffeine as the target chemical of the safety assessment, appears to be sufficiently conservative to protect human health. Our data also confirm that the NGRA framework passed the reality check and includes an acceptable (medium to high) level of confidence. This was strengthened by a traditional safety evaluation of caffeine on the basis of the available animal data (see step 9; OECD, 2020). Caffeine was shown to be less toxic compared to its most potent analogue theophylline which is not unexpected given the higher potency of theophylline compared to caffeine.

Funding body information

This work was funded through the Long Range Science Strategy (LRSS) programme of Cosmetics Europe (https://www.lrsscosmeticseurope.eu).

CRediT authorship contribution statement

Dagmar Bury: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Camilla Alexander-White:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Harvey J. Clewell:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Mark Cronin:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Bertrand Desprez:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Ann Detroyer:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Alina Efremenko:**

Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **James Firman:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Eric Hack:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Nicola J. Hewitt:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Gerry Kenna:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Martina Klaric:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Cathy Lester:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Catherine Mahony:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Gladys Ouedraogo:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Alicia Paini:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Andreas Schepky:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge Jane Rose from P&G as well as Gerhard J. Nohynek for reviewing parts of the manuscript.

References

Alexander-White, C., Mahony, C., Bury, D., Cronin, M., Dent, M., Hack, E., Hewitt, N., Kenna, G., Naciff, J., Ouedraogo, G., Shepky, A., 2021. A 10-step framework for use of read-across (RAX) in next generation risk assessment (NGRA) for cosmetics safety assessment. Manuscript submitted to. Regul. Toxicol. Pharmacol.

Arnaud, M.J., 2010. Pharmacokinetics and metabolism of natural methylxanthines in animal and man. Handb. Exp. Pharmacol. 200, pp33–91. https://doi.org/10.1007/978-3-642-13443-2_3. ISBN 978-3-642-13442-5. PMID 20859793.

Berggren, E., White, A., Ouedraogo, G., Paini, A., Richarz, A.N., Bois, F.Y., Exner, T., Leite, S., van Grunsven, L.A., Worth, A., Mahony, C., 2017. Ab initio chemical safety assessment: a workflow based on exposure considerations and non-animal methods. Comput. Toxicol. 4, 31–44.

Berthou, F., Ratanasavanh, D., Alix, D., Carlhant, D., Riche, C., Guillouzo, A., 1988. Caffeine and theophylline metabolism in newborn and adult human hepatocytes; comparison with adult rat hepatocytes. Biochem. Pharmacol. 37 (19), 3691–3700. https://doi.org/10.1016/0006-2952(88)90402-9.

Berthou, F., Ratanasavanh, D., Riche, C., Picart, D., Voirin, T., Guillouzo, A., 1989. Comparison of caffeine metabolism by slices, microsomes and hepatocyte cultures from adult human liver. Xenobiotica 19 (4), 401–417.

Bessems, J.G.M., Paini, A., Gajewska, M., Worth, A., 2017. The margin of internal exposure (MOIE) concept for dermal risk assessment based on oral toxicity data – a case study with caffeine. Toxicology 392, 119–129.

Bonati, M., Garattini, S., 1984. Interspecies Comparisons of Caffeine Disposition. In: Dews, P.B. (Ed.), Caffeine: Perspectives from Recent Research. Springer, Berlin, pp. 48–56.

Box, G.E.P., 1976. Science and statistics. J. Am. Stat. Assoc. 71 (356), 791–799. https://doi.org/10.1080/01621459.1976.10480949.

Brent, R.L., Christian, M.S., Diener, R.M., 2011. Evaluation of the reproductive and developmental risks of caffeine. Birth Defects Research (Part B) 92, 152–187.

Brown, R.P., Delp, M.D., Lindstedt, S.L., Rhomberg, L.R., Beliles, R.P., 1997. Physiological parameter values for physiologically based pharmacokinetic models. Toxicol. Ind. Health 13 (4), 407–484.

Campbell Jr., J.L., Clewell, R.A., Gentry, P.R., Andersen, M.E., Clewell III, H.J., 2012. Physiologically based pharmacokinetic/toxicokinetic modeling. In: Reisfeld, B., Mayeno, A.N. (Eds.), Computational Toxicology. Volume 1, vol. 929. Methods Mol Biol., pp. 439–499

Chavan, S., Judson, R., Patlewicz, G., Stark, R., Russell, P., Carmichael, P., 2017. Combining in Vivo, in Vitro and Toxicokinetics Data in Read-Across: A Case Study Using Caffeine. Poster, SOT meeting, Baltimore, MA, USA. March 12-16.

Cosmetics Ingredient Review (CIR), 2019. Safety assessment of methylxanthines as used in cosmetics. January 2019. https://online.personalcarecouncil.org/ctfa-static/online/lists/cir-pdfs/FR773.pdf.

Clewell, H.J., Lee, T., Carpenter, R.L., 1994. Sensitivity of physiologically-based pharmacokinetic models to variation in model parameters: methylene chloride. Risk Anal. 14, 521–531.

Clewell, H.J., Andersen, H.J., Blaauboer, B.J., 2008. On the incorporation of chemical-specific information in risk assessment. Toxicol. Lett. 180, 100–109.

Clewell, R.A., Clewell, H.J., 2008. Development and specification of physiologically based pharmacokinetic models for use in risk assessment. Regul. Toxicol. Pharmacol. 50, 129–143.

Comiskey, D., Api, A.M., Barratt, C., Daly, E.J., Ellis, G., Namara, C., O'Mahony, C., Robison, S.H., Safford, B., Smith, B., Tozer, S., 2015. Novel database for exposure to fragrance ingredients in cosmetics and personal care products. Regul. Toxicol. Pharmacol. 72 (3), 660–672.

Daly, J.W., Butts-Lamb, P., Padgett, W., 1983. Subclasses of adenosine receptors in the central nervous system: interaction with caffeine and related methylxanthines. Cell. Mol. Neurobiol. 3 (1), 69–80.

Denaro, C.P., Brown, C.R., Jacob III, P., Benowitz, N.L., 1991. Effects of caffeine with repeated dosing. Eur. J. Clin. Pharmacol. 40, 273–278.

Dias, M., Farinha, A., Faustino, E., Hadgraft, J., Pais, J., Toscano, C., 1999. Topical delivery of caffeine from some commercial formulations. Int. J. Pharm. 182, 41–47.

Doucet, O., Ferrero, L., Garcia, N., Zastrow, L., 1998. O/W emulsion and W/O/W multiple emulsion: physical characterization and Skin pharmacokinetic comparison in the delivery process of caffeine. Int. J. Cosmet. Sci. 20, 283–295.

ECHA, 2017. Read-across Assessment Framework (RAAF) ECHA-17-R-01-EN. https://doi.org/10.2823/61921.2, 978-92-9495-758-0.

ECHA/EFSA, 2018. Guidance for the Identification of Endocrine Disruptors in the Context of Regulations (EU) No 528/2012 and (EC) No 1107/2009. Prepublication version, 7 June 2018.

EFSA, 2015. Scientific Opinion on the safety of caffeine. EFSA Journal 13 (5), 4102.

EMA, 2018. Guideline on the reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-reporting-physiologically-based-pharmacokinetic-pbpk-modelling-simulation_en.pdf.

Fay, K.A., Villeneuve, D.L., Swintek, J., Edwards, S.W., Nelms, M.D., Blackwell, B.R., Ankley, G.T., 2018. Differentiating pathway-specific from nonspecific effects in high-throughput toxicity data: a foundation for prioritizing adverse outcome pathway development. Toxicol. Sci. 163 (2), 500–515. https://doi.org/10.1093/toxsci/kfy049.

Gajewska, M., Paini, A., Sala Benito, J.V., Burton, J., Worth, A., Urani, C., Briesen, H., Schramm, K.W., 2015. In vitro-to-in vivo correlation of the skin penetration, liver clearance and hepatotoxicity of caffeine. Food Chem. Toxicol. 75, 39–49.

Géniès, C., Jamin, E.L., Debrauwer, L., Zalko, D., Person, E.N., Eilstein, J., Grégoire, S., Schepky, A., Lange, D., Ellison, C., Roe, A., Salhi, S., Cubberley, R., Hewitt, N.J., Rothe, H., Klaric, M., Duplan, H., Jacques-Jamin, C., 2019. Comparison of the metabolism of 10 chemicals in human and pig skin explants. J. Appl. Toxicol. 39, 385–397.

Ginsberg, G., Hattis, D., Russ, A., Sonawane, B., 2004. Physiologically based pharmacokinetic (PBPK) modeling of caffeine and theophylline in neonates and adults: implications for assessing children's risks from environmental agents. J. Toxicol. Environ. Health, Part A 67, 297–329.

Gracia-Lor, E., Rousis, N.I., Zuccato, E., Bade, R., Baz-Lomba, J.A., Castrignanò, E., Causanilles, A., Hernández, F., Kasprzyk-Hordern, B., Kinyua, J., McCall, A.K., van Nuijs, A., Plósz, B.G., Ramin, P., Ryu, Y., Santos, M.M., Thomas, K., de Voogt, P., Yang, Z., Castiglioni, S., 2017. Estimation of caffeine intake from analysis of caffeine metabolites in wastewater. Sci. Total Environ. 609, 1582–1588. https://doi.org/10.1016/j.scitotenv.2017.07.258.

Hewitt, N.J., Grégoire, S., Cubberley, R., Duplan, H., Eilstein, J., Ellison, C., Lester, C., Fabian, E., Fernandez, J., Géniès, C., Jacques-Jamin, C., Klaric, M., Rothe, H., Sorrell, I., Lange, D., Schepky, A., 2020. Measurement of the penetration of 56

cosmetic relevant chemicals into and through human skin using a standardised protocol. J. Appl. Toxicol. 40 (3), 403–415.
Khera, K.S., 1985. Maternal toxicity: a possible etiological factor in embryo-fetal deaths and fetal malformations of rodent-rabbit species. Teratology 31, 129–153.
Klimisch, H.J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul. Toxicol. Pharmacol. 25 (1), 1–5.
Kolsek, K., Mavri, J., Sollner Dolenc, M., Gobec, S., Turk, S., 2014. Endocrine disruptome - an open source prediction tool for assessing endocrine disruption potential through nuclear receptor binding. J. Chem. Inf. Model. 54 (4), 1254–1267.
Kopecna, M., Macháček, M., Prchalová, E., Štěpánek, P., Drašar, P., Kotora, M., Vávrová, K., 2017. Dodecyl amino glucoside enhances transdermal and topical drug delivery via reversible interaction with skin barrier lipids. Pharm. Res. (N. Y.) 34, 640–653.
Lave, T., Dupin, S., Schmitt, C., Chou, R.C., Jaeck, D., Coassolo, P., 1997. Integration of in vitro data into allometric scaling to predict hepatic metabolic clearance in man: application to 10 extensively metabolised drugs. J. Pharmacol. Sci. 86 (5), 584–590.
Lelo, A., Birkett, D.J., Robson, R.A., Miners, J.O., 1986. Comparative pharmacokinetics of caffeine and its primary demethylated metabolites paraxanthine, theobromine and theophylline in man. Br. J. Clin. Pharmacol. 22, 177–182.
Livingston, E.H., Lee, S., 2001. Body surface area prediction in normal-weight and obese patients. Am. J. Physiol. Endocrinol. Metab. 281, E586–E591.
Luo, L., Lane, M.E., 2015. Topical and transdermal delivery of caffeine. Int. J. Pharmaceutics 490, 155–164.
Madden, J.C., Enoch, S.J., Paini, A., Cronin, M.T.D., 2020. A review of in silico tools as alternatives to animal testing: principles, resources and applications. Altern Lab Anim 48 (4), 146–172.
Monteiro, J.P., Alves, M.G., Oliveira, P.F., Silva, B.M., 2016. Structure-bioactivity relationships of methylxanthines: trying to make sense of all the promises and the drawbacks. Molecules 21, 974.
Monteiro, J.P., Alves, M.G., Oliveira, P.F., Silva, B.M., 2019. Pharmacological potential of methylxanthines: retrospective analysis and future expectations. Crit. Rev. Food Sci. Nutr. 59 (16), 2597–2625. https://doi.org/10.1080/10408398.2018.1461607.
Moxon, T.E., Li, H., Lee, M.Y., Piechota, P., Nicol, B., Pickles, J., Pendlington, R., Sorrell, I., Baltazar, M.T., 2020. Application of physiologically based kinetic (PBK) modelling in the next generation risk assessment of dermally applied consumer products. Toxicol. Vitro 63. https://doi.org/10.1016/j.tiv.2019.104746, 104746. PMID: 31837441.
Müller, C.E., Jacobson, K.A., 2011. Xanthines as adenosine receptor antagonists. Handb. Exp. Pharmacol. (200), 151–199. https://doi.org/10.1007/978-3-642-13443-2_6. PMID:20859796; PMCID: PMC3882893.
NTP, 1998. Tech. Rep. No. 473, Theophylline. National Toxicology Program, US. Department of Health and Human Services, NIH-Publication No, 98-3963; NTIS, PB99-113342).
Nyska, A., Herbert, R.A., Chan, P.C., Haseman, J.K., Hailey, J.R., 1998. Theophylline-induced mesenteric periarteritis in F344/N rats. Arch. Toxicol. 72 (Issue 11), 731–737.
OECD, 2001. Theophylline, CAS 58-55-9. SIDS Report.
OECD, 2002. Caffeine, CAS: 58-08-2. SIDS Initial Assessment Report for SIAM 14.
OECD, 2020. Case Study on the Use of Integrated Approaches for Testing and Assessment for Systemic Toxicity Arising from Cosmetic Exposure to Caffeine. Series on Testing and Assessment No, p. 321.
OECD, 2021. Guidance Document on the Characterisation, Validation and Reporting of PBK Models for Regulatory Purposes. Series on Testing and Assessment No. 331. ENV/CBC/MONO(2021)1. OECD Publishing, Paris.
Otberg, N., Patzelt, A., Rasulev, U., Hagemeister, T., Linscheid, M., Sinkgraven, R., Sterry, W., Lademann, J., 2008. The role of hair follicles in the percutaneous absorption of caffeine. Br. J. Clin. Pharmacol. 65, 488–492.
Paini, A., Leonard, J.A., Joossens, E., Bessems, J., Desalegn, A., Dorne, J.L., Gosling, J.P., Heringa, M.B., Klaric, M., Kliment, T., Kramer, N.I., Loizou, G., Louisse, J., Lumen, A., Madden, J.C., Patterson, E.A., Proença, S., Punt, A., Setzer, R.W., Suciu, N., Troutman, J., Yoon, M., Worth, A., Tan, Y.M., 2019. Next generation physiologically based kinetic (NG-PBK) models in support of regulatory decision making. Computational toxicology (Amsterdam, Netherlands) 9, 61–72. https://doi.org/10.1016/j.comtox.2018.11.002.
Pletz, J., Blakeman, S., Paini, A., Parissis, N., Worth, A., Andersson, A.M., Frederiksen, H., Sakhi, A.K., Thomsen, C., Bopp, S.K., 2020. Physiologically based kinetic (PBK) modelling and human biomonitoring data for mixture risk assessment. Environ. Int. 143 https://doi.org/10.1016/j.envint.2020.105978, 105978. Advance online publication.
Sachana, M., 2019. An international effort to promote the regulatory use of PBK models based on non-animal data. Computational Toxicology 11, 23–24. https://doi.org/10.1016/j.comtox.2019.01.002.
Safford, B., Api, A.M., Barratt, C., Comiskey, D., Daly, E.J., Ellis, G., McNamara, C., O'Mahony, C., Robison, S., Smith, B., Thomas, R., Tozer, S., 2015. Use of an aggregate exposure model to estimate consumer exposure to fragrance ingredients in personal care and cosmetic products. Regul. Toxicol. Pharmacol. 72 (3), 673–682.
SCCS, 2021. Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation. SCCS, 11th Revision. 30-31 March 2021. SCCS/1628/21.
Schmitt, W., 2008. General approach for the calculation of tissue to plasma partition coefficients. Toxicol. Vitro 22, 457–467.
Schuhmacher-Wolz, U., Voss, J.U., Schneider, K., 2017. Case study with natural substances on the various options to identify and categorise endocrine disruptors. J. Toxicol. Health 4 (2), 1–14. https://doi.org/10.7243/2056-3779-4-2.
Schultz, T.W., Amcoff, P., Berggren, E., Gautier, F., Klaric, M., Knight, D.J., Mahony, C., Schwarz, M., White, A., Cronin, M.T.D., 2015. A strategy for structuring and reporting a read-across prediction of toxicity. Regul. Toxicol. Pharmacol. 72, 586–601.
Shibata, M., Wachi, M., Kawaguchi, M., Kojima, J., Onodera, K., 2000. Teratogenic and foetal toxicity following intravenous theophylline administration in pregnant rabbits is related to maternal drug plasma levels. Methods Find. Exp. Clin. Pharmacol. 22, 101–107.
Shibata, Y., Takahashi, H., Chiba, M., Ishii, Y., 2002. Prediction of hepatic clearance and availability by cryopreserved human hepatocytes: an application of serum incubation method. Drug Metabol. Dispos. 30 (8), 892–896.
Smetanova, L., Stetinova, V., Kholova, D., Kvetina, J., Smetana, J., Svoboda, Z., 2009. Caco-2 cells and Biopharmaceutics Classification System (BCS) for prediction of transepithelial transport of xenobiotics (model drug: caffeine). Neuroendocrinol. Lett. 30 (Suppl. 1), 101–105.
Tarka Jr., S.M., 1982. The toxicology of cocoa and methylxanthines: a review of the literature. Crit. Rev. Toxicol. 9 (4), 275–312 cited in: Theocorp Holding Company LLC (2010). GRAS Exemption claim for Theobromine.
Theocorp Holding Company, L.L.C., 2010. GRAS Exemption Claim for Theobromine (3,7-dimethylxanthine), Summary of Data Concerning the Safety and GRAS Determination of Theobromine (3,7-dimethylxanthine) for Use as an Ingredient in Specified Foods.
Toner, F., Garcia, B., Roper, C., Madden, S., 2009. Effect of Repeat Application of [14C]-Caffeine on Percutaneous Absorption through Human Skin in Vitro. Poster, Occupational and Environmental Exposures of Skin to Chemicals. OEESC) meeting, Edinburgh.
Tozer, S.A., Kelly, S., O'Mahony, C., Daly, E.J., Nash, J.F., 2015. Aggregate exposure modelling of zinc pyrithione in rinse-off personal cleansing products using a person-orientated approach with market share refinement. Food Chem. Toxicol. 83, 103–110.
Tozer, S.A., O'Mahony, C., Hannah, J., O'Brien, J., Kelly, S., Kosemund-Meynen, K., Alexander-White, C., 2019. Aggregate exposure modelling of vitamin A from cosmetic products, diet and food supplements. Food Chem. Toxicol. 131 https://doi.org/10.1016/j.fct.2019.05.057, 110549.
Trauer, S., Patzelt, A., Otberg, N., Knorr, F., Rozycki, C., Balizs, G., Büttemeyer, R., Linscheid, M., Liebsch, M., Lademann, J., 2009. Permeation of topically applied caffeine through human skin – a comparison of in vivo and in vitro data. Br. J. Clin. Pharmacol. 68 (2), 181–186.
Troutman, J.A., Rick, D.L., Stuard, S.B., Fisher, J., Bartels, M.J., 2015. Development of a physiologically-based pharmacokinetic model of 2-phenoxyethanol and its metabolite phenoxyacetic acid in rats and humans to address toxicokinetic uncertainty in risk assessment. Regul. Toxicol. Pharmacol. 73, 530–543.
United States Environmental Protection Agency (USEPA), 1994. Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry. EPA/600/8-90/066F.
United States Environmental Protection Agency (USEPA), 2006. Approaches for the Application of Physiologically Based Pharmacokinetic (PBPK) Models and Supporting Data in Risk Assessment. National Center for Environmental Assessment, Washington, DC. EPA/600/R-05/043F. Available from: National Technical Information Service, Springfield, VA, and online at. http://epa.gov/ncea.
United States Environmental Protection Agency (USEPA), 2011. Recommended use of body weight3/4 as the default method in derivation of the oral reference dose. Risk Assessment Forum. EPA/100/R11/0001.
WHO, 2010. Characterization and Application of Physiologically Based Pharmacokinetic Models in Risk Assessment, vol. 9. IPCS Harmonization Project Document.
Wolde, T., 2014. Effects of caffeine on health and nutrition: a Review. Food Sci. Qual. Manag. 30, 59–65.
Wu, S., Blackburn, K., Amburgey, J., Jaworska, J., Federle, T., 2010. A framework for using structural, reactivity, metabolic and physicochemical similarity to evaluate the suitability of analogues for SAR-based toxicological assessments. Regul. Toxicol. Pharmacol. 56, 67–81.
York, R.G., Randall, J.L., Scott Jr., W.J., 1986. Teratogenicity of paraxanthine (1,7-dimethylxanthine) in C57BL/6J mice. Teratology 34, 279–282.
Zesch, A., Schaefer, H., Stüttgen, G., 1979. The quantitative distribution of percutaneously applied caffeine in the human skin. Arch. Dermatol. Res. 266 (3), 277–283.