# In silico prediction of Heterocyclic Aromatic Amines metabolism susceptible to form DNA adducts in humans

Victorien Delannée, Sophie Langouët, Anne Siegel, Nathalie Théret

Toxicology Letters 300 (2019) 18–30

Contents lists available at ScienceDirect

Toxicology Letters

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

# *In silico* prediction of Heterocyclic Aromatic Amines metabolism susceptible to form DNA adducts in humans

Victorien Delannée[a,b], Sophie Langouët[b], Anne Siegel[a], Nathalie Théret[b,*]

[a] Univ Rennes, Inria, CNRS, IRISA, F-35000 Rennes, France
[b] Univ Rennes, Inserm, EHESP, Irset – UMR_S1085, F-35043 Rennes, France

ARTICLE INFO

*Keywords:*
*In silico*
Xenobiotics
Metabolism
DNA adducts

ABSTRACT

Heterocyclic Aromatic Amines (HAAs) are environmental and food contaminants that are classified as probable or possible carcinogens by the International Agency for Research on Cancer. Thirty different HAAs have been identified. However the metabolism of only three of them have been fully characterized in human hepatocytes: AαC (2-amino-9H-pyrido[2,3-b]indole), MeIQx (2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline) and PhIP (2-amino-1-methyl-6-phenyl-imidazo[4,5-b]pyridine). In this study, we use an integrative approach to accurately predict the biotransformation of 30 HAAs into DNA reactive and non DNA reactive compounds. We first build predicted metabolites networks by iterating a knowledge-based expert system of prediction of metabolic reactions based on fingerprint similarities. Next, we combine several methods for predicting Sites Of Metabolism (SOM) in order to reduce the metabolite reaction graphs and to predict the metabolites reactive with DNA. We validate the method by comparing the experimental versus predicted data for the known AαC, MeIQx and PhIP metabolism. 28 of the 30 experimentally determined metabolites are well predicted and 9 of the 10 metabolites known to form DNA adducts are predicted with a high probability to be reactive with DNA. Applying our approach to the 27 unknown HAAs, we generate maps for the metabolic biotransformation of each HAA, including new metabolites with a high-predicted DNA reactivity, which can be further explored through an user-friendly and interactive web interface.

## 1. Introduction

Heterocyclic Aromatic Amines (HAAs) are food and environmental contaminants formed particularly in well-done cooked meats and during the burning of tobacco and diesel exhausts (Ni et al., 2008; Turesky and Le Marchand, 2011; Oz and Kaya, 2011). Based on toxicology studies, they have been classified as probable or possible human carcinogens (Group 2A and 2B) by the International Agency of Research on Cancer (N Authors Listed, 1987). Thirty HAAs have been identified (Pais and Knize, 2000; Ni et al., 2008; Turesky and Le Marchand, 2011). However the metabolism of only three has been fully characterized in human hepatocytes: PhIP (2-amino-1-methyl-6-phenyl-imidazo[4,5-*b*]pyridine) (Langouët et al., 2002), MeIQx (2-amino-3,8-dimethylimidazo[4,5-*f*]quinoxaline) (Langouët et al., 2001) and AαC (2-amino-9*H*-pyrido[2,3-*b*]indole) (Bellamri et al., 2017). HAAs metabolism occurs in two phases called biotransformation for phase I and conjugation for phase II. During Phase I, cytochromes P450 (CYP) activate the chemical substance by introducing a reactive and polar group. During phase II, Sulfotransferases (SULT), N-Acetytransferases (NAT) and UDP-Glucuronosyltransferases (UGT) (Turesky and Le Marchand, 2011) are major enzymes involved in conjugation of the activated metabolite by increasing its molecular weight and reducing its reactivity (Liska, 1998). In addition, Phase II can produce metabolites reactive with DNA leading to DNA adducts (Schut and Snyderwine, 1999) (Fig. 1). The lack of data about most of HAAs prompted us to develop *in silico* approaches for predicting large scale metabolic networks and the formation of DNA adducts.

While prediction of xenobiotic metabolism in humans is required to evaluate their toxicity and adverse effects, it remains a challenging task for pharmaceutical, cosmetics and food industries. Most of the existing approaches focus either on the prediction of metabolic reactions or on the prediction of sites of metabolism (SOM). The prediction of metabolic reactions is based on expert systems, where metabolites are predicted using dictionaries of biotransformation operators. These dictionaries are based on experimental data and contain biotransformation rules that change target fragments toward product fragments. The putative derivatives of a metabolite are computed by identifying all possible target fragments from the dictionaries appearing in the metabolite

* Corresponding author.
*E-mail address:* nathalie.theret@univ-rennes1.fr (N. Théret).

https://doi.org/10.1016/j.toxlet.2018.10.011
Received 17 July 2018; Received in revised form 2 October 2018; Accepted 8 October 2018
Available online 10 October 2018

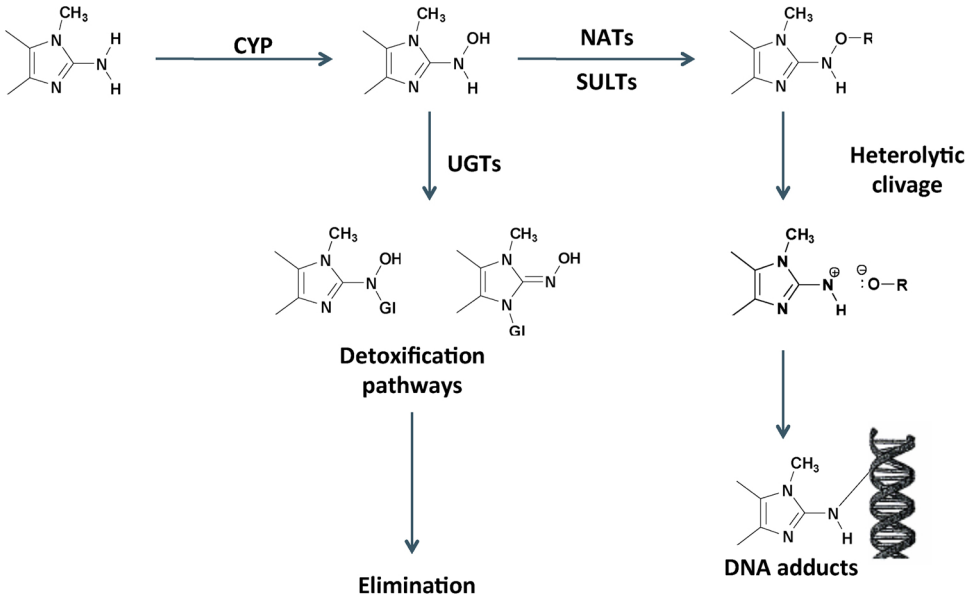



**Fig. 1. General biotransformation pathways of HAAs in the liver**. HAAs undergo hydroxylation mediated by CYPS and the intermediates can be either detoxified thanks to UGT enzymes or bioactivated by NAT and SULT enzymes leading to a compound reactive with DNA.

structure, and replacing them with their corresponding product fragments. Examples of state-of-the-art software solutions based on expert systems are MetabolExpert (Darvas, 1987), META (Klopman et al., 1994, 1997; Talafous et al., 1994) and METEOR (Button et al., 2003; Marchant et al., 2008). More recently, MetaPrint2D-React proposed a new approach based on data mining and statistical analysis (Adams, 2010). MetaPrint2D-React uses circular fingerprints of each atom of the queried molecule. Predicted metabolites are generated by applying biotransformation rules based on fingerprint similarity. One of the main drawbacks of the knowledge-based expert systems is the combinatorial explosion of predictions, leading to a high number of false positives (Bugrim et al., 2004). In order to improve precision while limiting the loss of sensitivity, cutoffs and probability levels are implemented. However, the precision remains low when a high sensitivity is needed (Piechota et al., 2013) and the precision improvement is associated with a decrease in sensitivity.

Besides the methods for predicting metabolic reactions, numerous tools for predicting sites of metabolism (SOM) have been developed to optimize metabolic properties of a chemical molecule during drug development process (see review Kirchmair et al., 2012). Among them, Xenosite (Zaretzki et al., 2013; Dang et al., 2016) and Way2Drug (Rudik et al., 2015) predict SOM for CYP and UGT. Due to the diversity of approaches for SOM prediction, some hybrid solutions have been proposed such as IMPACTS (Campagna-Slater et al., 2012), SMARTCYP (Rydberg et al., 2010) and Metasite (Cruciani et al., 2005).

In order to optimize the prediction of xenobiotics and drugs metabolism, some computational approaches combined predictions of biotransformation reactions and sites of metabolism. Tarcsay et al. (2010) used first the software predicting xenobiotic metabolites, MetabolExpert (Darvas, 1987) and next the docking program GLIDE (Friesner et al., 2004; Halgren et al., 2004) to increase the precision. Similarly, Piechota et al. (2013) developed a method combining metabolites prediction by Metaprint2D-React (Adams, 2010) and SOM prediction by SMARTCYP (Rydberg et al., 2010). Such hybrid approaches aim to reduce the number of over-predicted metabolites. However they remain unsatisfactory for predicting large maps of biotransformation metabolism either because of their low sensitivity or because of their poor scalability since docking-based methods are very time-consuming.

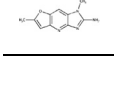
In this present work, we developed an original integrative method to accurately predict and study the different routes of transformation of HAAs into toxic and non-toxic compounds. Our approach pushes forward the hybrid approach of Piechota et al. (2013) by combining several methods to address both the issue of reducing over-predicted metabolites with expert-system methods and the issue of improving SOM identification. Our method consists first in over-approximating metabolic routes by an iterative use of Metaprint2D-React (Adams, 2010) leading to the production of exhaustive metabolic maps for each HAA. In order to overcome the issue of over-prediction while maintaining a high sensitivity and increasing the precision, we implement filtering methods based on SOM prediction by combining different tools: Metaprint2D-React (Adams, 2010), WhichCyp (Rostkowski et al., 2013), Way2Drug-SOMP (Rudik et al., 2015), Xenosite Metabolism 1.0 (Zaretzki et al., 2013), Xenosite UGT 2.0 (Dang et al., 2016) and Xenosite Reactivity 2.0 (Hughes et al., 2015, 2016). Finally, we used predictions from Xenosite Reactivity 2.0 (Hughes et al., 2015, 2016) to identify metabolites that are reactive with DNA. In order to combine these different tools, we introduced score-based criteria to accurately estimate enzyme specificity and DNA-reactivity. The tools were calibrated using manually curated compound databases independent from HAAs. The predictions of the method applied to three referenced HAAs demonstrate that combining different approaches and parameterizing them using manually curated knowledge leads of a high gain of sensitivity without any loss of precision. When applying our method to the 27 unknown HAA, we generated maps for the metabolic biotransformations of each HAA, including new metabolites with a high predicted DNA reactivity, which can be further explored through a user-friendly and interactive web interface.

## 2. Materials and methods

### 2.1. Software

To study HAA metabolism and predict the metabolites leading to DNA adducts, we implemented a workflow combining a set of software tools including Metaprint2D-React (Adams, 2010), WhichCyp (Rostkowski et al., 2013), Way2Drug-SOMP (Rudik et al., 2015), Xenosite Metabolism 1.0 (Zaretzki et al., 2013), Xenosite UGT 2.0 (Dang et al., 2016), and Xenosite Reactivity 2.0 (Hughes et al., 2015, 2016). Way2Drug and Xenosite are combined to potentiate SOMs identification and to overcome the possible lack of SOM recognition by using a

**Table 1**
HAAs dataset: List of the 30 Heterocyclic Aromatic Amines including AαC, MeIQx, and PhIP. They are classified in two categories that include Pyrolysis Heterocyclic Aromatic Amines and Aminoimidazoaren Heterocyclic Aromatic Amines depending on their formation temperature.

| Image | Name | IUPAC Name |
|---|---|---|
| *Pyrolysis Heterocylclic Aromatic Amines* | | |
|  | AαC | 9H-pyrido[2,3-b]indol-2-amine |
|  | MeAαC | 3-methyl-9H-pyrido[2,3-b]indol-2-amine |
|  | Glu-P-1 | 10-methyl-1,3,8-triazatricyclo[7.4.0.0$^{2,7}$] trideca-2(7),3,5,8,10,12-hexaen-4-amine |
|  | Glu-P-2 | 1,3,8-triazatricyclo[7.4.0.0$^{2,7}$]trideca-2(7),3,5,8,10,12-hexaen-4-amine |
|  | Tr-P-1 | 1,4-dimethyl-5H-pyrido[4,3-b]indol-3-amine |
|  | Tr-P-2 | 1-methyl-5H-pyrido[4,3-b]indol-3-amine |
|  | NorHarman | 9H-pyrido[3,4-b]indole |
|  | Harman | 1-methyl-9H-pyrido[3,4-b]indole |
|  | APNH | 4-{9H-pyrido[3,4-b]indol-9-yl}aniline |
|  | AMPNH | 2-methyl-4-9H-pyrido[3,4-b]indol-9-ylaniline |
|  | Phe-P-1 | 5-phenylpyridin-2-amine |
| *Aminoimidazoarene Heterocylclic Aromatic Amines* | | |
|  | IQ | 3-methyl-3H-imidazo[4,5-f]quinolin-2-amine |
|  | MeIQ | 3,4-dimethyl-3H-imidazo[4,5-f]quinolin-2-amine |
|  | IQx | 3-methyl-3H-imidazo[4,5-f]quinoxalin-2-amine |
|  | MeIQx | 3,8-dimethyl-3H-imidazo[4,5-f]quinoxalin-2-amine |
|  | 4-$CH_2$OH-8-MeIQx | {2-amino-3,8-dimethyl-3H-imidazo[4,5-f] quinoxalin-4- yl}methanol |
|  | 4,8-DiMeIQx | 3,4,8-trimethyl-3H-imidazo[4,5-f]quinoxalin-2-amine |
|  | 7,8-DiMeIQx | 3,7,8-trimethyl-3H-imidazo[4,5-f]quinoxalin-2-amine |
|  | 4,7,8-TriMeIQx | 3,4,7,8-tetramethyl-3H-imidazo[4,5-f] quinoxalin-2-amine |
|  | IQ[4,5-b] | 1-methyl-1H-imidazo[4,5-b]quinolin-2-amine |
|  | IgQx | 1-methyl-1H-imidazo[4,5-g]quinoxalin-2-amine |
|  | 7-MeIgQx | 1,7-dimethyl-1H-imidazo[4,5-g]quinoxalin-2-amine |
|  | 7,9-DiMeIgQx | 1,7,9-trimethyl-1H-imidazo[4,5-g]quinoxalin-2-amine |
|  | 6,7-DiMeIgQx | 1,6,7-trimethyl-1H-imidazo[4,5-g]quinoxalin-2-amine |
|  | 1,6-DMIP | 1,6-dimethyl-1H-1,3-benzodiazol-2-amine |
|  | 1,5,6-TMIP | 1,5,6-trimethyl-1H-1,3-benzodiazol-2-amine |
|  | 3,5,6-TMIP | 3,5,6-trimethyl-3H-imidazo[4,5-b]pyridin-2-amine |
|  | PhIP | 1-methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-amine |
|  | 4′-OH-PhIP | 4-{2-amino-1-methyl-1H-imidazo[4,5-b] pyridin-6-yl}phenol |
|  | IFP | 6,11-dimethyl-10-oxa-2,4,6-triazatricyclo [7.3.0.0$^{3,7}$]dodeca-1(9),2,4,7,11-pentaen-5-amine |

single method. The main lines of the workflow consist in generating a graph by the iterative use of Metaprint2D-React. This graph is next processed by software tools for SOM prediction (WhichCyp, Way2Drug SOMP, Xenosite Metabolism 1.0 and Xenosite UGT 2.0) in order to reduce over-predicted metabolites and to identify metabolites reactive with DNA (Xenosite Reactivity 2.0).

*Metaprint2D-React* (Adams, 2010) is a software predicting metabolic transformation in human, dog and rat models using 3 configuration parameters (Loose, Default, and Strict). We use Metaprint2D-React in an iterative way. Each considered HAA is converted to a SMILES (Simplified Molecular Input Line Entry Specification) string and used as input to Metaprint2D-React. The fingerprint matching parameters are initialized as default and the human model is selected. The metabolites predicted by the tool are used as new inputs in MetaPrint2D-React with the same parameters. We iterate this procedure in order to obtain two levels of metabolites (depth = 2), meaning that the initial HAA can undergo two consecutive transformations as described in the general scheme of HAA metabolism (Fig. 1). In addition, the MetaPrint2D-React tool is used to compute a normalized occurrence ratio score (NOR) that reports the frequency of a reported SOM in the metabolite database. The NOR score for a given atom in a query molecule is defined as the ratio between the number of atoms identified as a reaction centre sharing a similar chemical environment with the query atom in Metaprint2D-React database and the total number of atoms (identified both as reaction centre and non-reaction centre) with a similar chemical environment in Metaprint2D-React database. Therefore, each predicted metabolite is associated with a score ranging from 0 to 1, which corresponds by extrapolation to its formation frequency.

*WhichCyp* (Rostkowski et al., 2013) is a software predicting the selectivity for CYP1A2, CYP2C9, CYP2C19, CYP2D6 and CYP3A4. It predicts which isoenzyme either binds or metabolizes a molecule using simple yes/no classification models. All molecules predicted by Metaprint2D-React having potentially P450-mediated transformations are used as input to WhichCyp tool to predict which CYP isoform binds to the molecule.

*Xenosite P450 Metabolism 1.0* (Zaretzki et al., 2013) is a software predicting SOM of a molecule for 1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, 3A4 CYP isoforms. All molecules predicted by Metaprint2D-React having potentially P450-mediated transformations are used as input to Xenosite P450 Metabolism 1.0 tool. The software is used to identify SOM and allows filtering the P450-mediated reactions predicted by MetaPrint2D-React in addition to Way2Drug tool. Xenosite P450 Metabolism 1.0 computes a probability score varying between 0 and 1 (a high probability to be a SOM is characterized by a high score).

*Way2Drug SOMP* (Rudik et al., 2015) is a software predicting SOM of a molecule for 1A2, 2C9, 2C19, 2D6 and 3A4 CYP isoforms and UGT. All molecules predicted by Metaprint2D-React having potentially P450- or/and UGT-mediated transformations are used as input to Way2Drug SOMP tool. The software computes a probability score varying between 0 and 1 (a high probability to be a SOM is characterized by a high score), which determines the probability that a labeled atom is the SOM of the appropriate enzyme.

*Xenosite UGT 2.0* (Dang et al., 2016) is a software predicting SOM of a molecule for UGT. All molecules predicted by Metaprint2D-React having potentially UGT-mediated transformations are used as input to Xenosite UGT 2.0 tool. Xenosite UGT 2.0 is used to identify SOM associated with a probability score varying between 0 and 1 (a high probability to be a SOM is characterized by a high score).

*Xenosite Reactivity 2.0* (Hughes et al., 2015, 2016) is a software predicting both sites of reactivity (SOR) and molecular reactivity. Each molecule and its derivatives derived from a metabolite reactive with DNA are used as input to Xenosite Reactivity 2.0. The software predicts the reactivity with DNA of any metabolite thereby identifying which molecule can potentially lead to DNA adducts. Xenosite Reactivity 2.0 computes a probability score of reactivity varying between 0 and 1 (a high probability to be reactive is characterized by a high score).

### 2.2. Datasets

In addition to the dataset containing the 30 HAAs, three datasets were used to calibrate the workflow and validate the prediction.

The *HAAs dataset* contains the 30 HAAs including AαC, MeIQx, and PhIP (Ni et al., 2008; Turesky and Le Marchand, 2011; Oz and Kaya, 2011) (Table 1). They are classified in two categories that include Pyrolysis Heterocyclic Aromatic Amines and Aminoimidazoaren Heterocyclic Aromatic Amines depending on their formation temperature.

The *SOM calibration dataset* contains 18 molecules with aromatic amines that can undergo hydroxylation and/or glucuronidation reactions (Supplementary Material Table 4) (Beland, 1990; Castañeda-Acosta et al., 1999; Schut et al., 1984; Weisburger and Weisburger, 1958; N Authors Listed, 1989; Committee on Amines, 1981; Hammons et al., 1985; Parkinson and Ogilvie, 2010; Kadlubar et al., 1977; Orzechowski et al., 1992; Butler et al., 1989; Franklin, 1998; Frederick et al., 1985; Morton et al., 1980; Eiermann et al., 1998; Hiroi et al., 2002; Timbrell and Marrs, 2009; Alonen et al., 2005; Jiang et al., 2015; Baughman et al., 2009; Domínguez-Romero et al., 2012; Turesky, 2006), which are the two main classes of reactions involved in HAAs metabolism. These 18 molecules are involved in 59 reactions of 5 different types (C-Hydroxylation, N-Hydroxylation, C-Glucuronidation, N-Glucuronidation and O-Glucuronidation) occurring mostly on different aromatic amines. This dataset is used to calibrate the parameters of all software tools.

The *DNA adduct calibration dataset* contains 30 manually curated aromatic amines, that are known to form DNA adducts (Supplementary Material Table 5). This dataset is used to define different thresholds for Xenosite Reactivity 2.0 in order to classify metabolites according to their DNA reactivity.

The *validation dataset* contains 30 known derivatives of AαC, MeIQx, and PhIP (Supplementary Material Table 6). This dataset is used to

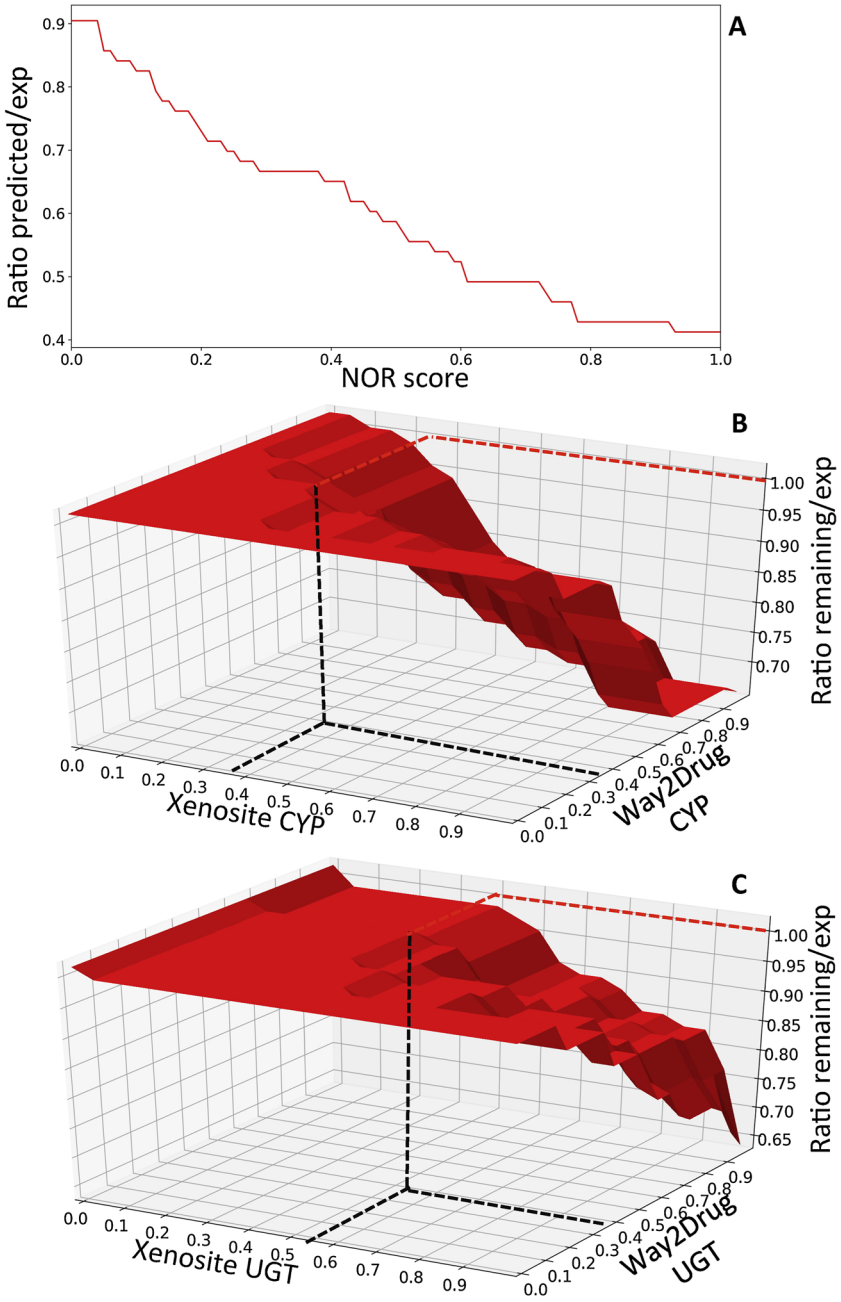



**Fig. 2. Evaluation of filtration thresholds**. (a) Analysis of the ratio between the number of predicted metabolites remaining after filtering and the number of experimentally determined metabolites, as a function of the Metaprint2d-React NOR Score. Each reaction and the corresponding metabolites of the SOM calibration dataset are predicted using basal configuration. (b) Analysis of the ratio between the number of predicted metabolites remaining after filtering and the number of experimentally determined metabolites as a function of XenositeP450 metabolism 1.0 and Way2Drug for CYP scores; (c) Analysis of the ratio between the number of predicted metabolites remaining after filtering and the number of experimentally determined metabolites as a function of XenositeUGT 2.0 and Way2Drug for UGT scores. In b and c, the method consists of maximizing the parameters without losing in sensitivity. The best set of parameters is the set where all metabolites present in the SOM calibration dataset are not filtered.

validate the full pipeline after its calibration with the different thresholds learned thanks to the other datasets. It allows estimating the sensitivity and the precision of the pipeline.

### 2.3. Tool parameters optimization

The calibration of the different thresholds of tools used the *SOM calibration dataset*, the *DNA adduct calibration dataset* and the *validation dataset*. To that goal, the Metaprint2D-React filtration procedure (based on a NOR-base threshold), the CYP filtration procedure (involving thresholds for Xenosite Metabolism 1.0 and Way2Drug) and the UGT filtration procedure (involving thresholds for Xenosite UGT 2.0 and Way2Drug) are applied to the datasets by parsing the complete range of thresholds available for the tools.

*Prediction of reactions with MetaPrint2D-React*. We used MetaPrint2D-React to compute the predicted products (and their associated NOR score) for all reactions that are known to produce a compound listed in *SOM calibration dataset*. Our computations permitted to predict 28 of the 30 known metabolites from A$\alpha$C, MeIQx and PhIP (Table 6). The two remaining derivatives 7-oxo-MeIQx and N-desmethyl-7-oxo-MeIQx are never predicted by MetaPrint2D-React and consequently will not influence the thresholds (this observation can be seen at a threshold of 0, which keep all metabolites). In order to calibrate MetaPrint2D-React, we choose a very low NOR score (0.04) as the maximum score to keep the 28 metabolites both identified by experiments and predicted by MetaPrint2D-React. Importantly half of other metabolites predicted for all HAAs have a NOR threshold below 0.5 emphasizing the huge sensitivity of the NOR score. In order to leverage the bias induced by the sensitivity of MetaPrint2D-React to the NOR score, we decided to keep a threshold of 0.04 to filter all outputs of MetaPrint2D-Reac.

*Prediction of SOM for UGT-mediated bioproducts*. The scores of the Way2Drug SOMP and the Xenosite UGT 2.0 software applied to each SOM of metabolites from *SOM calibration dataset* and associated with a UGT-mediated transformation were computed and plotted (see Fig. 2). This computation confirmed that no tool is sensitive enough to discriminate the experimentally determined SOMs (some have a score equal to 0 according to one software prediction). By contrast, this analysis demonstrated that for all SOMs, the scores for Xenosite UGT Metabolism 1.0 and Way2Drug SOMP were greater than 0.54 and 0.36, respectively. These parameters are the optimal ones to keep all the validated metabolites from *SOM calibration dataset*. Therefore, we considered that for any metabolite resulting from a UGT-mediated reaction, a site is a SOM for a glucoronidation if its Xenosite Metabolism 1.0's score is greater than 0.54, or its Way2Drug SOMP's score is greater than 0.36.

*Prediction of SOM for P450 isoform-mediated bioproducts*. The scores of the Way2Drug SOMP and the Xenosite P450 Metabolism 1.0 software applied to each SOM of metabolites from *SOM calibration dataset* and associated with a P450 isoform-mediated transformation predicted by WhichCyp, were computed and plotted (see Fig. 2). This computation confirmed that no tool is sensitive enough to discriminate the experimentally determined SOMs. By contrast, this analysis demonstrated that for all SOM, the scores for Xenosite Metabolism P450 1.0 and Way2Drug SOMP were greater than 0.39 and 0.35, respectively. These parameters are the optimal ones to keep all the metabolites of the *SOM calibration dataset*. Therefore, we considered that for any metabolite resulting from a P450-mediated reaction, a site is a SOM for hydroxylation if its Xenosite Metabolism 1.0 score is greater than 0.39, or its Way2Drug SOMP score is greater than 0.35.

*DNA reactivity classification* A similar approach based on the study of experimentally determined compounds was used to calibrate and evaluate the prediction of DNA adduct reactivity with Xenosite Reactivity 2.0. To that goal, the predicted DNA reactivity (in terms of probability) of a set of 30 manually-curated aromatic amines that are known to form DNA adducts was calculated. Our computations showed that 50% of the tested metabolites from *DNA adduct calibration dataset* have a probability between [0.85,1] to form DNA adducts, 20% have a probability which ranges between [0.7,0.85], 20% have a probability which ranges between [0.2,0.7] and 10% have a probability ranging between [0.05,0.2]. None of metabolites have a probability lower than 0.05. Based on these results, we considered the following criteria to classify the putative DNA-adduct reactivity of metabolites (genotoxicity): high probability of DNA-adduct reactivity (score greater than 0.85), medium probability (score between 0.7 and 0.85), low probability (score between 0.2 and 0.7), very-low probability (score between 0.05 and 0.2) and null-probability (score below 0.05). By extrapolation, we assume that the latter class contains non-genotoxic compounds.

*HAA-prediction evaluation* We evaluated the performance of our method to predict HAA derivatives at several steps of the workflow by computing the sensitivity (number of correctly predicted metabolites) and the precision scores with respect to the prediction of 30 known derivatives of A$\alpha$C, MeIQx, and PhIP. The *sensitivity* percentage was defined as the proportion of metabolites predicted to be produced from A$\alpha$C, MeIQx, and PhIP that are also experimentally determined (e.g., they belong to the *validation dataset*), that is to be $\frac{100[\{\text{predicted met.}\} \cap \{\text{exp. determined met.}\}]}{[\{\text{exp. determined met.}\}]}$. The *precision* percentage takes into account the number of over-predicted metabolites (not experimentally identified). It was defined to be equal to $\frac{100[\{\text{predicted met.}\} \cap \{\text{exp. determined met.}\}]}{[\{\text{predicted met.}\}]}$.

### 2.4. Web interface for map exploration

A web interface based on d3.js has been developed to visualize and manipulate the predicted metabolic pathways for the 30 HAAs. It enables the dynamic modification of the predicted graph either by selecting sets of reaction types or by displaying the reactivity of each metabolites. Thanks to this interface each predicted metabolite can also be independently studied by clicking on the metabolite of interest. The corresponding SMILES, and the predicted bio-transformations can be displayed by clicking on atoms. In addition, the final graph can also be exported in different formats such as JSON (containing all analysis information), or in image format (PDF, PNG, and svg). This tool is available at http://eppigraph.genouest.org/.

## 3. Results

### 3.1. Building graph of metabolic derivatives of HAA

In order to analyze HAAs metabolic pathways and predict the metabolic derivatives leading to DNA adducts, we developed a workflow combining 6 different tools (Metaprint2D-React, WhichCyp, Way2Drug SOMP, Xenosite Metabolism 1.0, Xenosite UGT 2.0 and Xenosite Reactivity 2.0). This workflow is composed of four steps depicted in Fig. 3.

- The first step is a fingerprint-based prediction. It consists in building a metabolic map by iterating the Metaprint2D-React tool twice.
- The second step is a knowledge-based reduction. It consists in selecting reactions which are known to be involved in HAAs metabolism (Hydroxylation, Glucoronidation, Acetylatin, Sulfation, Oxidation).
- The third step is a reduction based on site activities. It consists in filtering reactions catalyzed either by CYP isoforms (predicted by the WhichCyp tool) or by UGT. Reactions are conserved if they occur on atoms which are predicted to be sites of metabolism (SOM) either by the Way2Drug or the Xenosite tools according to criteria derived from the study of experimentally determined compounds (see methods). In addition, all reactions of the second generation of metabolites increasing the DNA reactivity (computed by Xenosite Reactivity 2.0) but occurring on atoms different from those

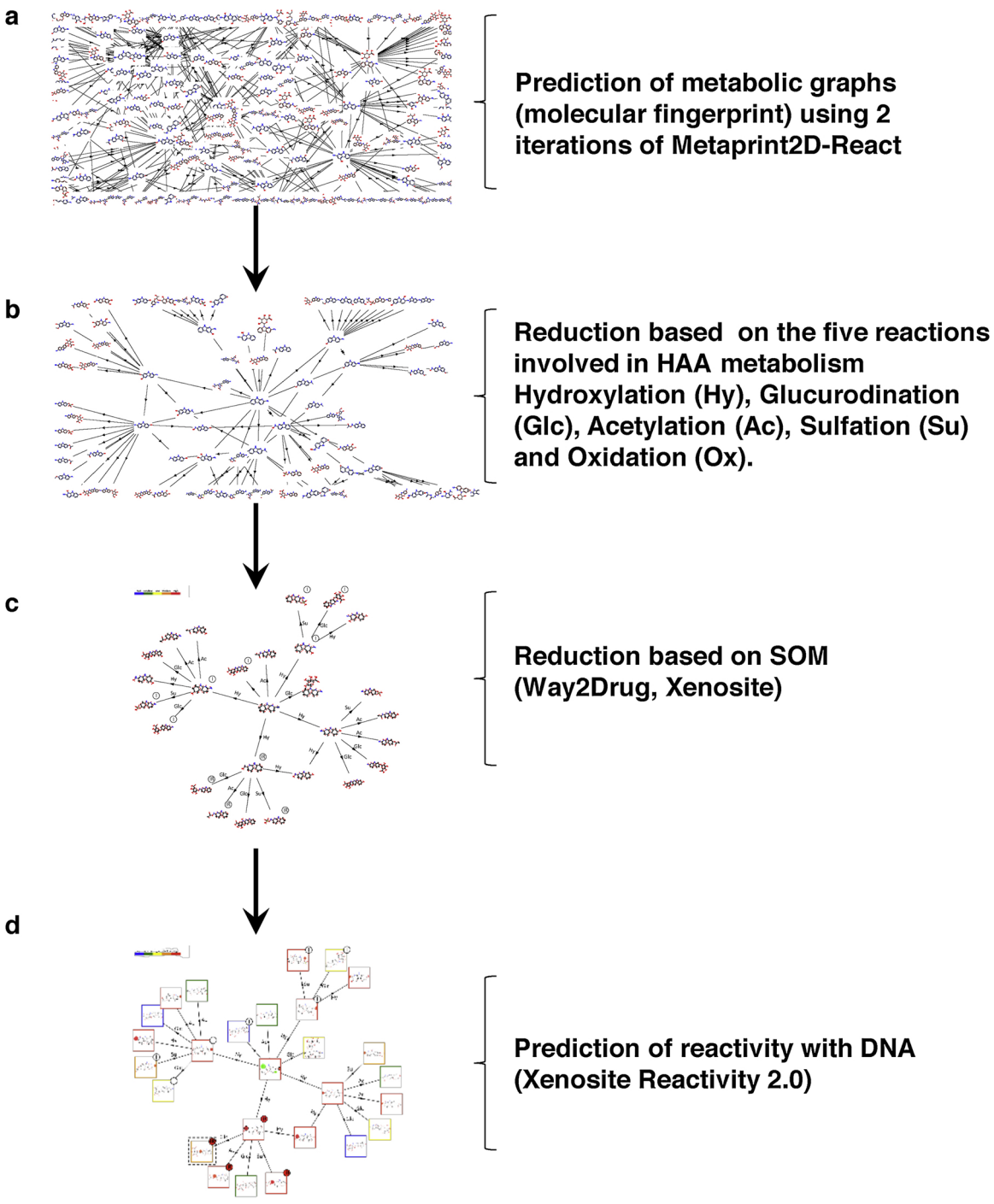



**Fig. 3. Schematic representation of the Workflow to predict metabolic pathways and metabolites reactive with DNA**. The workflow consists first to predict all set of reactions using two iterations of Metaprint2D-react which is calibrated to discard reactions occurring with a score lower than a previously learned NOR cutoff (a). Next, the predicted metabolic map is reduced through filtration steps including (b) the selection of Hydroxylation, Glucuronidation, Acetylation, Sulfation and Oxidation reactions and (c) the selection of reactions occurring on atoms which are predicted as SOM by Way2Drug and Xenosite Tools. The last step of the workflow relies on Xenosite Reactivity 2.0 (d) that keep the reactions increasing the DNA reactivity on atoms of molecules modified by the first generation. The output is a reduced graph that can be explored using Eppigraph (http://eppigraph.genouest.org/).

previously modified by a first generation reaction are deleted.

- The fourth step consists in classifying all the compounds of the final graph according to their DNA-reactivity.

At the end, given any input metabolite, the workflow produces a reduced metabolism graph of its derivatives enriched with the DNA adduct reactivity information.

### 3.2. Analysis of sensitivity and precision of the method

To validate the method, the workflow was first applied to three HAAs whose metabolic derivatives have been widely documented: A$\alpha$C, MeIQx and PhIP. The predicted metabolites, and their genotoxicity classification are described in Fig. 4–6. Next, the predicted metabolites were compared with the 30 experimentally determined metabolites and listed in Supplementary Table 5. As described in Table 2, MetaPrint2D-

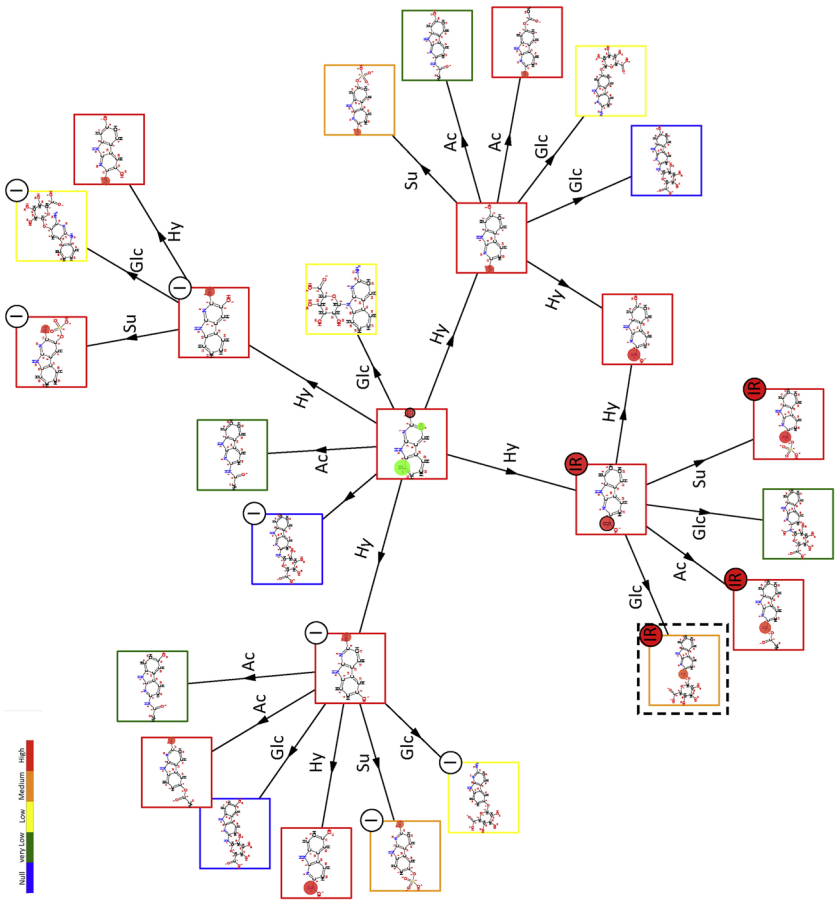



**Fig. 4. Graph of predicted metabolites for AαC**. Using the workflow, 26 derivatives were predicted for AαC. The genotoxicity of each metabolite is classified according to five levels of probability to form DNA adducts that are represented by colored frame: Null (blue), very Low (green), Low (yellow), Medium (orange), High (Red). The probability of each atom to bind DNA is indicated by a colored round:Low (Green), Medium (Yellow and Orange), High (Red). All experimentally determined metabolites are tagged with the letter I (Identified) and with the letter IR (as Identified and Reactive) when the metabolite is also known to be reactive with DNA. For AαC, 12 metabolites are predicted as highly reactive with DNA, 3 with a medium probability, 4 with a low probability, 4 with a very low probability and 3 with a null probability.

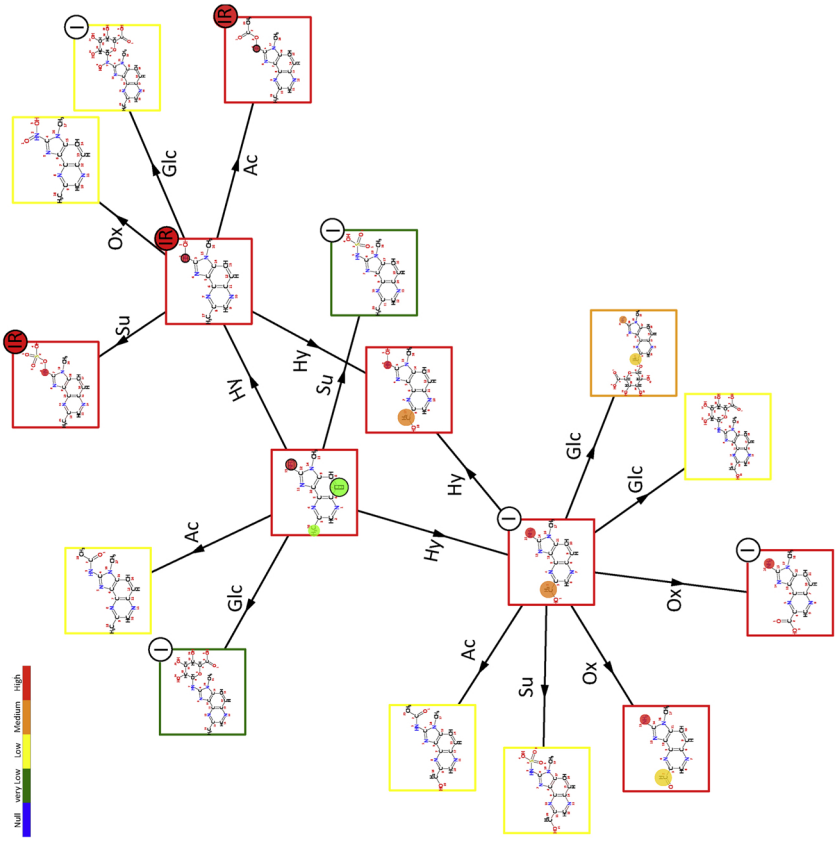



**Fig. 5. Graph of predicted metabolites for MeIQx**. Using the workflow, 16 derivatives were predicted for MeIQx. The genotoxicity of each metabolite is classified according to five levels of probability to form DNA adduct that are represented by colored frame: Null (blue), very Low (green), Low (yellow), Medium (orange), High (Red). The probability of each atom to bind DNA is indicated by a colored round:Low (Green), Medium (Yellow and Orange), High (Red). All experimentally determined metabolites are tagged with the letter I (Identified) and with the letter IR (as Identified and Reactive) when the metabolites is also known to be reactive with DNA. For MeIQx, 7 metabolites are predicted as highly reactive with DNA, 1 with a medium probability, 6 with a low probability, 2 with a very low probability and 0 with a null probability.

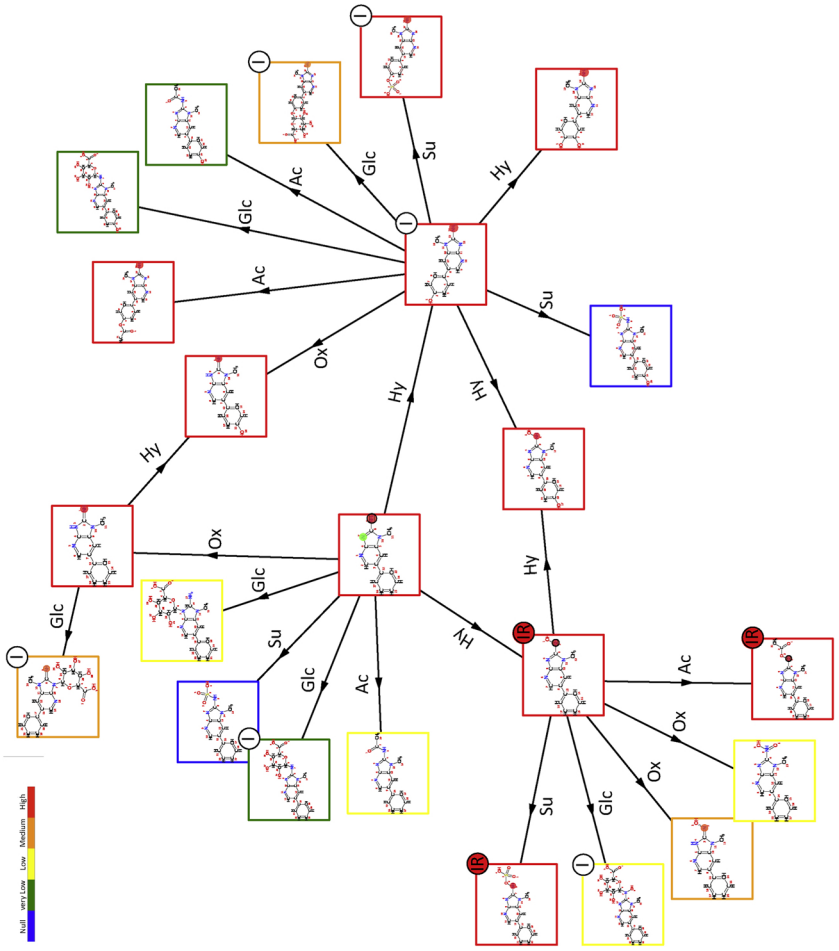



**Fig. 6. Graph of predicted metabolites for PhIP**. Using the workflow, 22 derivatives were predicted for PhIP. The genotoxicity of each metabolite is classified according to five levels of probability to form DNA adduct that are represented by colored frame: Null (blue), very Low (green), Low (yellow), Medium (orange), High (Red). The probability of each atom to bind DNA is indicated by a colored round: Low (Green), Medium (Yellow and Orange), High (Red). All experimentally determined metabolites are tagged with the letter I (Identified) and with the letter IR (as Identified and Reactive) when the metabolites is also known to be reactive with DNA. For PhIP, 10 metabolites are predicted as highly reactive with DNA, 3 with a medium probability, 4 with a low probability, 3 with a very low probability and 2 with a null probability.

**Table 2**

**Sensitivity and precision of the workflow**. The workflow was applied to A*α*C, MeIQx and PhIP and the predicted metabolites were compared to the 30 experimentally determined metabolites as derivatives of these three HAAs. The steps of the workflow include the iterative use of Metaprint2D-React, the filtering based on the five reactions occurring in HAAs metabolism (Hydroxylation, Glucuronidation, Acetylation, Sulfation and Oxidation) and the filtering based on CYP selectivity, SOM for CYP and UGT, and DNA-Reactivity. The *sensitivity* percentage is defined as the proportion of metabolites predicted to be produced from A*α*C, MeIQx, and PhIP that are also experimentally determined. The *precision* percentage takes into account the number of over-predicted metabolites (not experimentally determined).

| | Step of the workflow | Global | A*α*C | MeIQx | PhIP |
|---|---|---|---|---|---|
| | *1 iteration* | | | | |
| | Prediction of metabolism with 1 iteration of MetaPrint2D-React | 22% (11/50) | 22% (4/18) | 40% (4/10) | 14% (3/22) |
| Precision | Selection of Hydroxylation, Glucuronidation, Acetylation, Sulfation and Oxidation reactions | 50% (11/22) | 50% (4/8) | 80% (4/5) | 33% (3/9) |
| | Filtering based on CYP selectivity, SOM for CYP and UGT, DNA-Reactivity | 58% (11/19) | 57% (4/7) | 80% (4/5) | 43% (3/7) |
| | Prediction of metabolism with 1 iterations of MetaPrint2D-React | 85% (11/13) | 100% (4/4) | 67% (4/6) | 100% (3/3) |
| Sensitivity | Selection of Hydroxylation, Glucuronidation, Acetylation, Sulfation and Oxidation reactions | 85% (11/13) | 100% (4/4) | 67% (4/6) | 100% (3/3) |
| | Filtering based on CYP selectivity, SOM for CYP and UGT, DNA-Reactivity | 85% (11/13) | 100% (4/4) | 67% (4/6) | 100% (3/3) |
| | *2 iterations* | | | | |
| | Prediction of metabolism with 2 iterations of MetaPrint2D-React | 4% (28/786) | 4% (11/310) | 7% (8/107) | 2% (9/369) |
| Precision | Selection of Hydroxylation, Glucuronidation, Acetylation, Sulfation and Oxidation reactions | 23% (28/124) | 28% (11/40) | 38% (8/21) | 14% (9/63) |
| | Filtering based on CYP selectivity, SOM for CYP and UGT, DNA-Reactivity | 44% (28/64) | 42% (11/26) | 50% (8/16) | 41% (9/22) |
| | Prediction of metabolism with 2 iterations of MetaPrint2D-React | 93% (28/30) | 100% (11/11) | 80% (8/10) | 100% (9/9) |
| Sensitivity | Selection of Hydroxylation, Glucuronidation, Acetylation, Sulfation and Oxidation reactions | 93% (28/30) | 100% (11/11) | 80% (8/10) | 100% (9/9) |
| | Filtering based on CYP selectivity, SOM for CYP and UGT, DNA-Reactivity | 93% (28/30) | 100% (11/11) | 80% (8/10) | 100% (9/9) |

React identified 50 metabolites using one iteration and 786 metabolites using two iterations for the three HAAs A*α*C, MeIQx and PhIP (Global Column). Of note a single iteration of MetaPrint2D-React only predicted 11 over the 13 known first generation derivatives of the three HAAs while no metabolite of the second generation was predicted. This observation supports the iterative use of MetaPrint2D-react in order to predict the complete metabolism of molecules.

To evaluate the efficiency of the filtering procedures, the sensitivity and precision scores were calculated for each step of the workflow including the prediction of CYP selectivity, the prediction of SOM for CYP and UGT and the prediction of DNA-reactivity. We also implemented a filtering procedure to select the metabolites produced by the five different types of reaction (Hydroxylation, Glucuronidation, Acetylation, Sulfation and Oxidation) known to occur in the HAAs metabolism and

catalyzed by CYP, UGT, NAT and SULT enzymes (Bellamri et al., 2017; Langouët et al., 2002; Langouët et al., 2001). We observed that the filtration steps reduced the number of derivative metabolites of AαC, MeIQx, and PhIP from 786 to 64 (reduction rate of 81%) with an increased precision from 4 to 44% and a sensitivity of 93%. Importantly, the 28 known derivatives of these three HAAs predicted by Meta-Print2D-react were conserved by the filtering procedures suggesting that reduction steps discarded overpredicted metabolites without loss of precision. Note that the selection of the five reactions Hydroxylation, Glucuronidation, Acetylation, Sulfation and Oxidation permitted to reduce the prediction from 786 metabolites to 124 with an increased precision score from 4 to 23%. This observation demonstrated that focusing on major reactions of HAAs metabolism was very powerful and that integrating biology knowledge greatly improved precision of the method. When compared with data from Piechota et al., our workflow succeeded in increasing the precision without modifying the sensitivity. By changing the cutoffs implemented in Meteor tool, Piechota et al., predicted metabolites for phase I reactions with an increased precision from 3.4 to 10.7% but a loss of sensitivity from 85.2 to 78.2%.

To evaluate the prediction for genotoxicity, we next compared the compounds predicted to have a high DNA reactivity (identified by a red circle in Fig. 4–6) with the experimentally determined metabolites for AαC, MeIQx and PhIP. We defined the level of genotoxity of a HAA as the ratio of derivative metabolites having a high DNA-reactivity with respect to the total number of metabolites in the graph. As shown in the Table 3, the workflow predicted 7, 10 and 12 reactive metabolites for MeIQx, PhIP and AαC, respectively. Among these reactive metabolites 3, 3 and 4 have been already experimentally determined for MeIQx, PhIP and AαC, respectively (see Supplementary Table 5). Based on this data, the levels of genotoxicity for predicted metabolites were 43, 45 and 46% for AαC, MeIQx and PhIP respectively. These results are very close to the levels of genotoxicity calculated for experimentally determined metabolites (36, 38, 33%) suggesting the efficiency of our workflow.

Of note, N-sulfonyl-HAA and N-acetoxy-HAA that were previously described as reactive with DNA (Schut and Snyderwine, 1999) were predicted to have a high DNA reactivity using the workflow. In addition all metabolites previously identified as reactive with DNA after an N-hydroxylation reaction for AαC, MeIQx and PhIP (Turesky et al., 1991; Cai et al., 2016) were also correctly predicted with a high DNA reactivity. We further predicted that the metabolite produced after a glucuronidation reaction in AαC graph (Fig. 4, black dotted lines) is reactive with DNA thereby confirming experimental data recently published by Cai et al. (2016), Bellamri et al. (2017). However, the identification of this compound using purification methods from liver tissue (Bellamri et al., 2017) suggest stability in vivo. The medium probability of the prediction of reactivity might be associated with these two antagonistic features including the reactivity with DNA and the relative stability of the compound. While conjugation reactions mediated by UGT enzymes are often associated with detoxification pathways, these new data suggest potential genotoxicity of metabolite derivatives after glucuronidation reactions.

Together, our data demonstrate that the workflow predicts metabolites that have been already experimentally determined, demonstrating the efficacy of the workflow. In addition the analysis shows that the implementation of filtration considerably reduces the number of overpredicted metabolites (increased precision) without loss of experimentally determined metabolites (conserved sensibilities).

**Table 3**
**Classification of HAAs according to their predicted genotoxicity**. The level of genotoxicity was measured as the ratio between the number of reactive metabolites and the total number of metabolites in the predicted graph. For AαC, MeIQx and PhIP, the level of genotoxicity was also calculated for the experimentally determined metabolites (right columns).The predicted metabolites with high level of genotoxicity at the bottom of the table are expected to be potentially higher risk of mutagenicity/carcinogenicity.

| HAA | Predicted metabolites | | | Experimentally determined metabolites | | |
|---|---|---|---|---|---|---|
| | Number of DNA-reactive metabolites | Total number of metabolites | Level of genotoxicity | Number of DNA-reactive metabolites | Total number of metabolites | Level of genotoxicity |
| AMPNH | 0 | 1 | 0% | | | |
| APNH | 0 | 1 | 0% | | | |
| Harman | 0 | 0 | 0% | | | |
| NorHarman | 0 | 0 | 0% | | | |
| 1,5,6-TMIP | 3 | 8 | 37% | | | |
| 1,6-DMIP | 3 | 8 | 37% | | | |
| 4,7,8-TriMeIQx | 3 | 8 | 37% | | | |
| 6,7-DiMeIgQx | 3 | 8 | 37% | | | |
| 7,8-DiMeIQx | 3 | 8 | 37% | | | |
| IgQx | 3 | 8 | 37% | | | |
| IQ | 3 | 8 | 37% | | | |
| IQx | 3 | 8 | 37% | | | |
| MeIQ | 3 | 8 | 37% | | | |
| 3,5,6-TMIP | 4 | 10 | 40% | | | |
| 4′-OH-PhIP | 17 | 39 | 43% | | | |
| 4,8-DiMeIQx | 7 | 16 | 43% | | | |
| 7,9-DiMeIgQx | 7 | 16 | 43% | | | |
| 7-MeIgQx | 7 | 16 | 43% | | | |
| IFP | 4 | 10 | 40% | | | |
| MeIQx | 7 | 16 | 43% | 3 | 8 | 38% |
| PhIP | 10 | 22 | 45% | 3 | 9 | 33% |
| AαC | 12 | 26 | 46% | 4 | 11 | 36% |
| 4-CH2OH-8-MeIQx | 7 | 15 | 46% | | | |
| IQ[4,5-b] | 12 | 25 | 48% | | | |
| TrP1 | 4 | 8 | 50% | | | |
| GluP1 | 3 | 6 | 50% | | | |
| GluP2 | 3 | 6 | 50% | | | |
| TrP2 | 6 | 11 | 54% | | | |
| MeAαC | 12 | 22 | 54% | | | |
| PheP1 | 5 | 9 | 55% | | | |

### 3.3. Predicting the graphs of metabolism for 30 HAAs provides new insights on their genotoxicity and identifies candidates for the generation of DNA adducts

The workflow was applied to the 27 other HAAs to predict their graphs of metabolic derivatives and evaluate their genotoxicity. All predicted graphs enriched with the predicted DNA-reactivity of each metabolite are available online (http://eppigraph.genouest.org/). Raw graphs generated by MetaPrint2D-React can be easily explored using an intuitive user interface that permits to filter the graphs according to reactions and to export data files. Comparison of the predicted metabolites reactive with DNA permitted to classify all HAAs according to their level of genotoxicity (Table 3).

When applied to all HAAs, the workflow revealed a high heterogeneity in graph size since the number of identified metabolites ranged from 1 to 39 (Table 3). When considering AMPNH, APNH, Harman and NorHarman, the reduced graphs contain only 0 or 1 metabolite and did not permit to calculate the level of genotoxicity. While MetaPrint2D-React predicted more derivatives of these HAAs, they were all removed from metabolic graphs by the filtering procedure. Note that neither Harman nor NorHarman have an exocyclic amine group classically involved in biotransformation thereby preventing derivatives identification. In accordance with this, no carcinogenicity has been experimentally determined for these two HAAs (Totsuka et al., 2006). The remaining 26 HAAs including AαC, MeIQx and PhIP showed a tight distribution with a level of genotoxicity ranging from 37 to 55%. Interestingly, the six HAAs with highest levels of genotoxicity (TrP1, TrP2, GluP1, GluP2, MeαC and Phe-P-1) are structurally closed to AαC, a high carcinogen metabolite (Cai et al., 2016) and might be regarded as relevant genotoxic candidates. These predictions are in agreement with data from the International Agency for Research on Cancer which classified TrP1, TrP2, GluP2 and MeαC as probable carcinogens (class 2b). In contrast, HAAs characterized by a lower predicted genotoxicity belong to the Aminoimidazoarene Heterocylclic Aromatic Amines family suggesting that this family has a lower carcinogenicity than the pyrolytic HAAs.

In order to evaluate the impact of the 5 major reactions that include Hydroxylation, Glucuronidation, Acetylation, Sulfation and Oxidation reactions on the formation of metabolites reactive with DNA, we classified the reactions and their combination according to the DNA reactivity of the metabolites induced. As shown in Fig. 7, five classes were defined including null, very low, low, medium and high DNA reactivity. The metabolites resulting from the combinations of hydroxylation (Hy) and acetylation (Ac) or from hydroxylation (Hy) and sulfation (Su) have a high probability to be reactive with DNA. This observation is consistent with biological experiments from Cai et al. (2016), Schut and Snyderwine (1999) which showed that these chains of reactions are sources of labile intermediate metabolites with a high DNA reactivity thereby identifying bioactivation pathways. Similarly, the chains of reactions combining either hydroxylation and hydroxylation (Hy-Hy) in the graphs of 12 HAAs or hydroxylation and oxidation (Hy-Ox) in the graphs of 20 HAAs resulted in metabolites with a high probability to be reactive with DNA. Such chains of reactions contribute to putative novel bioactivation pathways leading to DNA adducts.

Importantly, such comparative analyses suggest that most of HAAs glucuronate conjugates have a low probability to be reactive with DNA since the chain of reactions involving glucuronidation (Glc) is mainly associated to the formation of metabolites with very low or low probabilities to be reactive with DNA (Fig. 7). While glucuronidation has been recently involved in potential bioactivation pathways for AαC metabolism (Cai et al., 2016), this reaction appeared less implicated in genotoxicity pathway when compared with other reactions. As reviewed in Kaivosaari et al. (2011), glucuronidation is firstly described as a reaction for the elimination of xenobiotics. In the present study, we showed that the chain of reactions combining hydroxylation and glucuronidation led to metabolites mostly involved in detoxification pathways but did not exclude the formation of potential bioactivated metabolites as demonstrated for AαC. Together our predictive analyses suggest that combination of hydroxylation and glucuronidation reactions might be involved both in bioactivation and detoxification processes. Based on this new hypothesis, the genotoxocity of other glucuronid derivatives predicted for 4,8-DiMeIQx, 7,9-DiMeIgQx, 7-MeIgQx, IQ[4,5-b], MeαC, MeIQx and PhIP might be reevaluated.

## 4. Discussion

The present work is the first large-scale study to predict HAAs metabolites and their capacity to form DNA adducts. Using an integrative computational approach, we predicted the derivatives of 30 HAAs and the chains of reactions leading to reactive metabolites.

The workflow that we developed is based on the iterative use of MetaPrint2D-React to predict a raw graph containing a large number of metabolites. The sensitivity of the method is confirmed by the identification of the 28 over 30 known metabolites of AαC, MeIQx and PhIP. In addition, this step of prediction permits to identify chains of reactions which cannot experimentally be detected due to the high lability of intermediates. Such an example is 1-methyl-6-phenyl-1H,2H,3H-imidazo[4,5-b]pyridin-2-imine that is predicted as an intermediate required for the generation of PhIP-N3-Gluc (Fig. 6).

By contrast, predicting a large graph of reactions is often associated with over-prediction and increase in background noise. To overcome this limitation, docking methods and predictions based on SOM are widely used to prioritize and classify metabolite predictions. Such approaches aim to combine precision (i.e. the capability not to over-predict metabolites) with the prediction of several generations of metabolites (Kirchmair et al., 2012). However, docking approaches are time and resources consuming and require information about isoforms involved in the xenobiotics metabolism (Tarcsay et al., 2010). In addition, methods based on SOM prediction loose in sensitivity while gaining in precision (Piechota et al., 2013).

In order, to address the over-prediction problem, our workflow combines several tools to filter the predicted metabolites including two different tools for SOM prediction. The rule used to combine these methods relies on the estimation of individual scores from a set of metabolites having properties close to that of HAAs. This process increases the precision by reducing the number of false positive metabolites. As exemplified in supplementary Fig. 8, the combination of Way2Drug and Xenosite tools enables a better identification of SOMs. In this case, Way2Drug tool failed to identify the NH2 group from AαC as a SOM for CYP1A2 and CYP2D6 while the two isoenzymes were predicted to metabolize AαC by WhichCYP tool. By contrast, Xenosite Metabolism 1.0 identified this SOM with a high probability. Note that the two other SOM predicted by Way2Drug are confirmed by Xenosite Metabolism 1.0 but with a low score.

Unlike Piechota et al. (2013) who used iterative approach to predict four metabolite generations, we performed predictions over two generations since reactive molecules are mostly produced within the second generation (Cai et al., 2016; Schut and Snyderwine, 1999). In addition, we selected the five major reactions involved in HAAs metabolism to make predictions more reliable. While Piechota et al. used cutoffs to increase the prediction to the detriment of sensitivity, we optimized the filtering steps on the basis of biological knowledge without loss of sensitivity. Such an approach is more powerful for generalizing the workflow to other components.

Using this filtration procedure, we predicted all experimentally determined metabolites and new derivatives for AαC, MeIQx, and PhIP. Among the new candidates, we cannot exclude false-positive namely because of the absence of filtration steps for reactions catalyzed by SULT and NAT enzymes. Therefore, the metabolites produced by SULT- or NAT-mediated reactions might be considered with parsimony and require validation using additional methods such as docking, new tools based on SOM prediction, or data-mining with an associated statistical

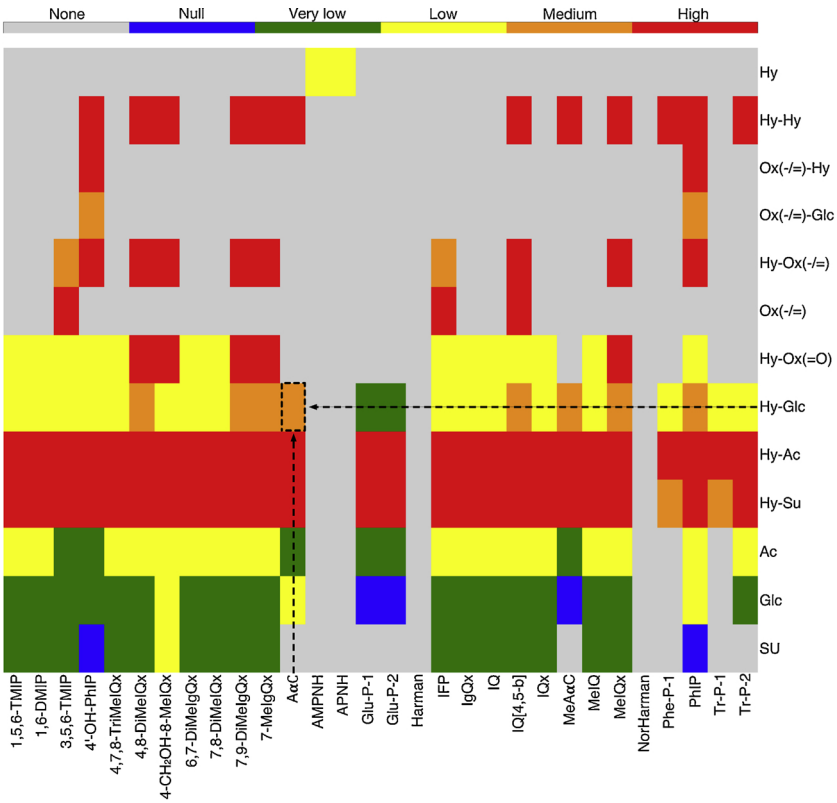



**Fig. 7. Clustering of HAAs according to DNA reactivity of metabolites resulting from combinations of Hydroxylation (Hy), Glucuronidation (Glc), Acetylation (Ac), Sulfation (Su) and Oxidation (Ox) reactions.** The graphs of metabolites were computed for all HAAs that were further clustered according to the occurrence of chains of one or two reactions present in at least two graphs. For each selected chain of reactions appearing in the graph of a given HAA, the set of metabolites derived from the combination of reactions is computed, as well as their predicted DNA-reactivity probabilities. Six levels of probability to form DNA adduct are defined: very Low (green), Low (yellow), Medium (orange), High (Red). "None" (grey) means that the chain of reactions does not appear in the graph of metabolism of the considered HAA. "null" (blue) means that the corresponding metabolite is considered as detoxified. The dotted line box delineated the pathways producing the glucuronate conjugates reactive with DNA from A*α*C and experimentally determined by Cai et al. (2016). Note, that two different oxidation are depicted: the oxidation of a single bound to a double bound (−/=) and the addition of an Oxygen link by a double bound (=O).

score in the manner of Metaprint2D.

An important hypothesis coming from our analysis is that some predicted metabolites have never been experimentally determined because of either their instability that does not permit to capture them in studied samples or the limitations of current technical methods that do not allow the detection of low quantity of metabolites. In support of this hypothesis, we noticed that none of PhIP intermediates predicted by our approach have been experimentally determined up to now. This is consistent with the fact that molecules reactive with DNA such as the N-Hydroxylated intermediates are known to be very labile and hardly measurable (Turesky, 1990). By contrast, metabolites with a lower reactivity with DNA are more stable and therefore more easily detected. According to this hypothesis, the O-glucuronate conjugate of A*α*C that we predicted with a low reactivity to DNA might be more stable than the N-acetoxy and N-sulfonyloxy derivatives (Cai et al., 2016; Schut and Snyderwine, 1999).

A major result of the present work is the comparative analysis of metabolic networks predicted for the 30 HAAs. We provide a database storing the predicted metabolic graphs of all HAAs through an interactive interface EPPIGRAPH that facilitates exploration and analysis. In addition, we show that the classification of HAAs according to the predicted score of genotoxicity is correlated to their distribution in the two major classes of HAAs. While the formation of all HAAs mainly results from heating condition of meats or combustion of tobacco, the structure of HAAs depends on temperature. Jagerstad et al. 1998 defined the thermic HAAs, IQ-type HAAs (imidazoquinoline or imidazoquinoxaline) or aminoimidazoazaarenes that are formed at temperatures between 100 and 300 degree C and the pyrolytic HAAs or non-IQ-type HAAs that are formed at high-temperature (> 300 degrees). Our workflow classified thermic HAAs as lower genotoxic agents (level of genotoxicity from 37 to 45%) and pyrolytic HAAs as higher genotoxic agents (from 46 to 55%). Such data are in agreement with literature (Gibis, 2016; Turesky and Le Marchand, 2011) and demonstrate the robustness of the workflow thereby providing identification of potential new genotoxic derivatives.

Combined with the prediction of reactivity, our approach provides hints on the nature of metabolites (intermediates or final). As an important perspective, the prediction of graphs of metabolic derivatives will provide good candidates for subtle mass spectrometry identification based on *a priori* knowledge of the chemical formula of the metabolic candidates.

To conclude, we developed a new efficient workflow to build metabolic networks and to predict metabolites reactive with DNA. We provided the first metabolite predictive maps for all HAAs family and identified potential new reactive metabolites. This paves the way to new experimental research towards a better and subtle estimation of risks of xenobiotics in human health.

## Acknowledgements

This work was supported by Institut National de la Santé et de la Recherche Médicale (Inserm), University of Rennes 1, la Ligue contre le Cancer du Grand Ouest, PNREST Anses, Cancer TMOI AVIESAN, [2013/1/166]. VD was a recipient of Région Bretagne.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.toxlet.2018.10.011.

## References

Adams, S.E., 2010. Molecular Similarity and Xenobiotic Metabolism, Thesis. University of Cambridge. https://www.repository.cam.ac.uk/handle/1810/225225.

Alonen, A., Aitio, O., Hakala, K., Luukkanen, L., Finel, M., Kostiainen, R., 2005. Biosynthesis of dobutamine monoglucuronides and glucuronidation of dobutamine by recombinant human Udp-glucuronosyltransferases. Drug Metab. Dispos. 33 (5), 657–663. https://doi.org/10.1124/dmd.104.002139. http://dmd.aspetjournals.org/content/33/5/657.

Baughman, T.M., Talarico, C.L., Soglia, J.R., 2009. Evaluation of the metabolism of propranolol by linear ion trap technology in mouse, rat, dog, monkey, and human cryopreserved hepatocytes. Rapid Commun. Mass Spectrom. 23 (14), 2146–2150.

https://doi.org/10.1002/rcm.4130.
Beland, F.A., 1990. Chemical Carcinogenesis and Mutagenesis I. Springer. http://www.springer.com/br/book/9783642747779.
Bellamri, M., Le Hegarat, L., Turesky, R.J., Langouët, S., 2017. Metabolism of the tobacco carcinogen 2-amino-9h-pyrido[2,3-b]indole (AαC) in primary human hepatocytes. Chem. Res. Toxicol. 30 (2), 657–668. https://doi.org/10.1021/acs.chemrestox.6b00394.
Bugrim, A., Nikolskaya, T., Nikolsky, Y., 2004. Early prediction of drug metabolism and toxicity: systems biology approach and modeling. Drug Discov. Today 9 (3), 127–135.
Butler, M.A., Guengerich, F.P., Kadlubar, F.F., 1989. Metabolic oxidation of the carcinogens 4-aminobiphenyl and 4,4'-methylenebis(2-chloroaniline) by human hepatic microsomes and by purified rat hepatic cytochrome P-450 monooxygenases. Cancer Res. 49 (1), 25–31. http://cancerres.aacrjournals.org/content/49/1/25.
Button, W.G., Judson, P.N., Long, A., Vessey, J.D., 2003. Using absolute and relative reasoning in the prediction of the potential metabolism of xenobiotics. J. Chem. Inf. Comput. Sci. 43 (5), 1371–1377. https://doi.org/10.1021/ci0202739.
Cai, T., Yao, L., Turesky, R.J., 2016. Bioactivation of heterocyclic aromatic amines by UDP glucuronosyltransferases. Chem. Res. Toxicol. 29 (5), 879–891. https://doi.org/10.1021/acs.chemrestox.6b00046.
Campagna-Slater, V., Pottel, J., Therrien, E., Cantin, L.-D., Moitessier, N., 2012. Development of a computational tool to rival experts in the prediction of sites of metabolism of xenobiotics by p450s. J. Chem. Inf. Model. 52 (9), 2471–2483. https://doi.org/10.1021/ci3003073.
Castañeda-Acosta, J., Bounds, P.L., Winston, G.W., 1999. Microsomal deacetylation of ring-hydroxylated 2-(acetylamino)fluorene isomers: effect of ring position and molecular mechanics considerations. J. Biochem. Mol. Toxicol. 13 (5), 279–286.
Committee on Amines, Clayson, D.B., 1981. Aromatic Amines: An Assessment of the Biological and Environmental Effects. National Academies google-Books-ID: slArAAAAYAAJ.
Cruciani, G., Carosati, E., De Boeck, B., Ethirajulu, K., Mackie, C., Howe, T., Vianello, R., 2005. MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. J. Med. Chem. 48 (22), 6970–6979. https://doi.org/10.1021/jm050529c.
Dang, N.L., Hughes, T.B., Krishnamurthy, V., Swamidass, S.J., 2016. A simple model predicts UGT-mediated metabolism. Bioinformatics (Oxford, England) 32 (20), 3183–3189. https://doi.org/10.1093/bioinformatics/btw350.
Darvas, F., 1987. Metabolexpert: an expert system for predicting metabolism of substances. In: Kaiser, K.L.E. (Ed.), QSAR in Environmental Toxicology – II. Springer Netherlands, pp. 71–81. https://doi.org/10.1007/978-94-009-3937-0_7.
Domínguez-Romero, J.C., García-Reyes, J.F., Martínez-Romero, R., Berton, P., Martínez-Lara, E., Del Moral-Leal, M.L., Molina-Díaz, A., 2013. Combined data mining strategy for the systematic identification of sport drug metabolites in urine by liquid chromatography time-of-flight mass spectrometry. Anal. Chim. Acta 761, 1–10. https://doi.org/10.1016/j.aca.2012.11.049.
Eiermann, B., Edlund, P.O., Tjernberg, A., Dalén, P., Dahl, M.-L., Bertilsson, L., 1998. 1- and 3-hydroxylations, in addition to 4-hydroxylation, of debrisoquine are catalyzed by cytochrome P450 2d6 in humans. Drug Metab. Dispos. 26 (11), 1096–1101. http://dmd.aspetjournals.org/content/26/11/1096.
Franklin, N., 1998. The N-glucuronidation of xenobiotics. An aspet-supported symposium held at the 1996 faseb meeting in Washington, DC. Drug Metab. Dispos.: Biol. Fate Chem. 26 (9), 829.
Frederick, C.B., Weis, C.C., Flammang, T.J., Martin, C.N., Kadlubar, F.F., 1985. Hepatic N-oxidation, acetyl-transfer and DNA-binding of the acetylated metabolites of the carcinogen, benzidine. Carcinogenesis 6 (7), 959–965.
Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P., Shenkin, P.S., 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 47 (7), 1739–1749. https://doi.org/10.1021/jm0306430.
Gibis, Monika, 2016. Heterocyclic aromatic amines in cooked meat products: causes, formation, occurrence, and risk assessment. Compr. Rev. Food Sci. Food Saf. 15 (2), 269–302.
Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T., Banks, J.L., 2004. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J. Med. Chem. 47 (7), 1750–1759. https://doi.org/10.1021/jm030644s.
Hammons, G.J., Guengerich, F.P., Weis, C.C., Beland, F.A., Kadlubar, F.F., 1985. Metabolic oxidation of carcinogenic arylamines by rat, dog, and human hepatic microsomes and by purified flavin-containing and cytochrome P-450 monooxygenases. Cancer Res. 45 (8), 3578–3585. http://cancerres.aacrjournals.org/content/45/8/3578.
Hiroi, T., Chow, T., Imaoka, S., Funae, Y., 2002. Catalytic specificity of CYP2d isoforms in rat and human. Drug Metab. Dispos. 30 (9), 970–976. https://doi.org/10.1124/dmd.30.9.970. http://dmd.aspetjournals.org/content/30/9/970.
Hughes, T.B., Miller, G.P., Swamidass, S.J., 2015. Site of reactivity models predict molecular reactivity of diverse chemicals with glutathione. Chem. Res. Toxicol. 28 (4), 797–809. https://doi.org/10.1021/acs.chemrestox.5b00017.
Hughes, T.B., Dang, N.L., Miller, G.P., Swamidass, S.J., 2016. Modeling reactivity to biological macromolecules with a deep multitask network. ACS Cent. Sci. 2 (8), 529–537. https://doi.org/10.1021/acscentsci.6b00162.
Jiang, L., Liang, S.-C., Wang, C., Ge, G.-B., Huo, X.-K., Qi, X.-Y., Deng, S., Liu, K.-X., Ma, X.-C., 2015. Identifying and applying a highly selective probe to simultaneously determine the O-glucuronidation activity of human UGT1a3 and UGT1a4. Sci. Rep. 5https://doi.org/10.1038/srep09627. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC096/4401.
Kadlubar, F.F., Miller, J.A., Miller, E.C., 1977. Hepatic microsomal N-glucuronidation and nucleic acid binding of N-hydroxy arylamines in relation to urinary bladder carcinogenesis. Cancer Res. 37 (3), 805–814. http://cancerres.aacrjournals.org/content/37/3/805.
Kaivosaari, S., Finel, M., Koskinen, M., 2011. N-glucuronidation of drugs and other xenobiotics by human and animal UDP-glucuronosyltransferases. Xenobiotica 41 (8), 652–669.
Kirchmair, J., Williamson, M.J., Tyzack, J.D., Tan, L., Bond, P.J., Bender, A., Glen, R.C., 2012. Computational prediction of metabolism: sites, products, SAR, P450 enzyme dynamics, and mechanisms. J. Chem. Inf. Model. 52 (3), 617–648.
Klopman, G., Dimayuga, M., Talafous, J., 1994. META. 1. A program for the evaluation of metabolic transformation of chemicals. J. Chem. Inf. Comput. Sci. 34 (6), 1320–1325.
Klopman, G., Tu, M., Talafous, J., 1997. META. 3. A genetic algorithm for metabolic transform priorities optimization. J. Chem. Inf. Comput. Sci. 37 (2), 329–334.
Langouët, S., Welti, D.H., Kerriguy, N., Fay, L.B., Huynh-Ba, T., Markovic, J., Guengerich, F.P., Guillouzo, A., Turesky, R.J., 2001. Metabolism of 2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline in human hepatocytes: 2-amino-3-methylimidazo[4,5-f]quinoxaline-8-carboxylic acid is a major detoxification pathway catalyzed by cytochrome P450 1a2. Chem. Res. Toxicol. 14 (2), 211–221.
Langouët, S., Paehler, A., Welti, D.H., Kerriguy, N., Guillouzo, A., Turesky, R.J., 2002. Differential metabolism of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine in rat and human hepatocytes. Carcinogenesis 23 (1), 115–122.
Liska, D.J., 1998. The detoxification enzyme systems. Altern. Med. Rev.: J. Clin. Ther. 3 (3), 187–198.
Marchant, C.A., Briggs, K.A., Long, A., 2008. In silico tools for sharing data and knowledge on toxicity and metabolism: derek for windows, meteor, and vatic. Toxicol. Mech. Methods 18 (2–3), 177–187. https://doi.org/10.1080/15376510701857320.
Morton, K.C., Beland, F.A., Evans, F.E., Fullerton, N.F., Kadlubar, F.F., 1980. Metabolic activation of N-hydroxy-N,N'-diacetylbenzidine by hepatic sulfotransferase. Cancer Res. 40 (3), 751–757. http://cancerres.aacrjournals.org/content/40/3/751.
N Authors Listed, 1987. Overall evaluations of carcinogenicity: an updating of IARC Monographs volumes 1 to 42. IARC Monogr. Eval. Carcinog. Risks Hum. Suppl. 7, 1–440.
N Authors Listed, 1989. 2-Nitrofluorene, IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, vol. 46. pp. 277–289.
Ni, W., McNaughton, L., LeMaster, D.M., Sinha, R., Turesky, R.J., 2008. Quantitation of 13 heterocyclic aromatic amines in cooked beef, pork, and chicken by liquid chromatography-electrospray ionization/tandem mass spectrometry. J. Agric. Food Chem. 56 (1), 68–78. https://doi.org/10.1021/jf072461a.
Orzechowski, A., Schrenk, D., Bock, K.W., 1992. Metabolism of 1- and 2-naphthylamine in isolated rat hepatocytes. Carcinogenesis 13 (12), 2227–2232.
Oz, F., Kaya, M., 2011. Heterocyclic aromatic amines in meat. J. Food Process. Preserv. 35 (6), 739–753. https://doi.org/10.1111/j.1745-4549.2011.00524.x/abstract.
Pais, P., Knize, M.G., 2000. Chromatographic and related techniques for the determination of aromatic heterocyclic amines in foods. J. Chromatogr. B Biomed. Sci. Appl. 747 (1–2), 139–169.
Parkinson, A., Ogilvie, B.W., 2010. Biotransformation of Xenobiotics – Casarett & Doull's Essentials of Toxicology, 2e – AccessPharmacy. McGraw-Hill Medical, McGraw-Hill Global Education Holdings (Chapter 6). http://accesspharmacy.mhmedical.com/Content.aspx?bookId=449&sectionId=39910772.
Piechota, P., Cronin, M.T.D., Hewitt, M., Madden, J.C., 2013. Pragmatic approaches to using computational methods to predict xenobiotic metabolism. J. Chem. Inf. Model. 53 (6), 1282–1293. https://doi.org/10.1021/ci400050v.
Rostkowski, M., Spjuth, O., Rydberg, P., 2013. WhichCyp: prediction of cytochromes P450 inhibition. Bioinformatics 29 (16), 2051–2052. https://doi.org/10.1093/bioinformatics/btt325. https://academic.oup.com/bioinformatics/article/29/16/2051/199965/WhichCyp-prediction-of-cytochromes-P450-inhibition.
Rudik, A., Dmitriev, A., Lagunin, A., Filimonov, D., Poroikov, V., 2015. SOMP: web server for in silico prediction of sites of metabolism for drug-like compounds. Bioinformatics (Oxford, England) 31 (12), 2046–2048. https://doi.org/10.1093/bioinformatics/btv087.
Rydberg, P., Gloriam, D.E., Zaretzki, J., Breneman, C., Olsen, L., 2010. SMARTCyp: a 2d method for prediction of cytochrome P450-mediated drug metabolism. ACS Med. Chem. Lett. 1 (3), 96–100. https://doi.org/10.1021/ml100016x.
Schut, H.A., Snyderwine, E.G., 1999. DNA adducts of heterocyclic amine food mutagens: implications for mutagenesis and carcinogenesis. Carcinogenesis 20 (3), 353–368.
Schut, H.A., Daniel, F.B., Schenck, K.M., Loeb, T.R., Stoner, G.D., 1984. Metabolism and DNA adduct formation of 2-acetylaminofluorene by bladder explants from human, dog, monkey, hamster and rat. Carcinogenesis 5 (10), 1287–1292.
Talafous, J., Sayre, L.M., Mieyal, J.J., Klopman, G., 1994. META. 2. A dictionary model of mammalian xenobiotic metabolism. J. Chem. Inf. Comput. Sci. 34 (6), 1326–1333.
Tarcsay, A., Kiss, R., Keseru, G.M., 2010. Site of metabolism prediction on cytochrome P450 2c9: a knowledge-based docking approach. J. Comput.-Aided Mol. Des. 24 (5), 399–408. https://doi.org/10.1007/s10822-010-9347-3.
Timbrell, J.A., Marrs, T.C., 2009. Biotransformation of Xenobiotics. General, Applied and Systems Toxicology. John Wiley & Sons, Ltd., pp. 1. https://doi.org/10.1002/9780470744307.gat004/abstract.
Totsuka, T., Nishigaki, R., Sugimura, T., Wakabayashi, K., 2006. The possible involvement of mutagenic and carcinogenic heteroyclic amines in human cancer. In: Skog, K., Alexander, J. (Eds.), Acrylamide and Other Hazardous Compounds in Heat-Treated Foods. Woodhead Publisher, Boca Raton, FL, pp. 296–515 Carcinogenesis.
Turesky, R.J., Le Marchand, L., 2011. Metabolism and biomarkers of heterocyclic aromatic amines in molecular epidemiology studies: lessons learned from aromatic amines. Chem. Res. Toxicol. 24 (8), 1169–1214. https://doi.org/10.1021/tx200135s. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3156293/.
Turesky, R.J., Lang, N.P., Butler, M.A., Teitel, C.H., Kadlubar, F.F., 1991. Metabolic

activation of carcinogenic heterocyclic aromatic amines by human liver and colon. Carcinogenesis 12 (10), 1839–1845.
Turesky, R.J., 1990. Metabolism and biodisposition of heterocyclic amines. Prog. Clin. Biol. Res. 347, 39–53.
Turesky, R.J., 2007. Formation and biochemistry of carcinogenic heterocyclic aromatic amines in cooked meats. Toxicol. Lett. 168 (3), 219–227. https://doi.org/10.1016/j.toxlet.2006.10.018.
Weisburger, E.K., Weisburger, J.H., 1958. Chemistry, carcinogenicity, and metabolism of 2-fluorenamine and related compounds. Adv. Cancer Res. 5, 331–431.
Zaretzki, J., Matlock, M., Swamidass, S.J., 2013. XenoSite: accurately predicting CYP-mediated sites of metabolism with neural networks. J. Chem. Inf. Model. 53 (12), 3373–3383. https://doi.org/10.1021/ci400518g.