



HAL
open science

Candidate Molecule Selection Based on In Silico Predicted ADMET Properties of 12 Indenoindole Derivatives

Nathalie Guragossian, Gustavo Jabor Gozzi, Bruno Fouillet, Marie-Paule Gustin, Raphaël Terreux, Zouhair Bouaziz, Christelle Marminon, Joachim Jose, Attilio Di Pietro, Markku Pasanen, et al.

► **To cite this version:**

Nathalie Guragossian, Gustavo Jabor Gozzi, Bruno Fouillet, Marie-Paule Gustin, Raphaël Terreux, et al.. Candidate Molecule Selection Based on In Silico Predicted ADMET Properties of 12 Indenoindole Derivatives. *Chemical Informatics*, 2016, 2 ((2)6), 10.21767/2470-6973.100020 . hal-01683399

HAL Id: hal-01683399

<https://hal.science/hal-01683399v1>

Submitted on 1 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/311557665>

Candidate Molecule Selection Based on In Silico Predicted ADMET Properties of 12 Indenoindole Derivatives

Article · November 2016

DOI: 10.21767/2470-6973.100020

CITATIONS

0

READS

67

11 authors, including:



Zouhair Bouaziz

Claude Bernard University Lyon 1

68 PUBLICATIONS 288 CITATIONS

[SEE PROFILE](#)



Joachim Jose

Westfalian Wilhelms-University of Münster

180 PUBLICATIONS 2,451 CITATIONS

[SEE PROFILE](#)



Attilio Di Pietro

Institute for the Biology and Chemistry of Pro...

188 PUBLICATIONS 4,740 CITATIONS

[SEE PROFILE](#)



Markku Pasanen

University of Eastern Finland

155 PUBLICATIONS 4,618 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Toxicity of waste water chemicals [View project](#)



Site specific glycoconjugation-Well define conjugation [View project](#)

DOI: 10.21767/2470-6973.100020

Candidate Molecule Selection Based on *In Silico* Predicted ADMET Properties of 12 Indenoindole Derivatives

Nathalie Guragossian¹,
Gustavo Jabor Gozzi²,
Bruno Fouillet³,
Marie-Paule Gustin⁴,
Raphaël Terreux⁵,
Zouhair Bouaziz¹,
Christelle Marminon¹,
Joachim Jose⁶,
Attilio Di Pietro²,
Markku Pasanen⁷ and
Marc Le Borgne¹

Abstract

For considering future *in vivo* assays, it is necessary to investigate pharmacokinetic and toxicity profile of new chemical entities to select the best candidate(s) for further evaluations. Physicochemical parameters and ADMET (Absorption, Distribution, Metabolism, Elimination and Toxicity) properties of 12 indenoindole derivatives – identified as potent inhibitors of the ABCG2 protein - were predicted *in silico* with the Molinspiration and the ACD/Percepta softwares. The evaluation of mutagenicity and carcinogenicity was achieved by using the QSAR Toolbox software. Based on the exercise, i) two phenolic derivatives should not be metabolically activated by CYP enzymes according to the QSAR Toolbox software leading to a lower mutagenic risk, ii) compounds 2b, 2c could be excluded from further studies because of clastogenic risks and again compound 2c for a relatively low oral bioavailability, iii) one compound for its blood toxicity and five because for their pulmonary toxicity. Finally, six out of the 12 derivatives (1a, 1b, 2a, 2d, 2e and 2g), were predicted, in terms of ADMET properties, to be good candidates for further *in vivo* investigations.

Keywords: ABCG2 inhibitors; Indenoindoles; ADMET; *In silico* predictions

Received: October 17, 2016; **Accepted:** November 04, 2016; **Published:** November 12, 2016

Background

A major problem of anticancer chemotherapy is multidrug resistance, often related to overexpression of multidrug ATP-binding cassette (ABC) transporters, which reduce the intracellular concentration of drugs below their cytotoxic threshold. Three among the 48 human ABC transporters are recognized to be associated with a low prognostic in cancer patients treated by chemotherapy: ABCB1/Pgp (P-glycoprotein) [1], ABCC1/MRP1 (multidrug resistance protein 1) [2], and the more recently identified ABCG2/BCRP (breast cancer resistance protein) [3-5]. The latter is a “half-transporter” of 655 aminoacids, containing one cytosolic nucleotide-binding domain, and one transmembrane domain with six-helical spans, which homodimerizes to be functional. It was found to be abundant in many types of cancer [6].

One of the strategies aimed at eliminating chemoresistant cancer cells is to use inhibitors of the multidrug ABC transporters in order to sensitize tumor growth to anticancer drug cytotoxicity. The first

- 1 Faculté de Pharmacie - ISPB, EA 4446 Bioactive Molecules and Medicinal Chemistry, Université de Lyon, Université Claude Bernard Lyon 1, Lyon Cedex 8, France
- 2 Institute of Protein Biology and Chemistry, MMSB, UMR 5086 CNRS, University of Lyon, Lyon, France
- 3 ISPBL, Laboratoire de Toxicologie, Faculté de Pharmacie, Université Lyon 1, Lyon, France
- 4 Emerging Pathogens Laboratory - Fondation Mérieux, Centre International de Recherche en Infectiologie (CIRI) Inserm U1111, CNRS UMR5308, ENS de Lyon, Lyon, France
- 5 Pôle Rhône-Alpes de Bioinformatique - Lyon Gerland (PRABI-LG), LBTI UMR 5305 CNRS UCBL, Université Lyon 1, Lyon, France
- 6 Institute of Pharmaceutical and Medicinal Chemistry, Pharma Campus, Westfälische Wilhelms - University Münster, Münster, Germany
- 7 School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland

Corresponding author: Marc Le Borgne

✉ marc.le-borgne@univ-lyon1.fr

Faculté de Pharmacie - ISPB, EA 4446
Bioactive Molecules and Medicinal
Chemistry, Université de Lyon, Université

specific ABCG2 inhibitor, of natural origin, was fumitremorgin C (FTC) which unfortunately displayed serious neurotoxicity [7]. Synthetic derivatives were developed, resulting in a highly potent inhibitor, Ko143, which however still retained some residual toxicity [8]. Screening of different classes of flavonoids identified interesting inhibitors such as hydrophobic flavones and acridones [9,10], chromones [11], and different types of chalcones [12], some of the inhibitors being active *in vivo* in mice models [13,14].

Novel ABCG2-selective inhibitors were recently developed as a series of ketonic indenoindoles after appropriate substitutions such as the replacement of isopropyl by phenethyl at *N*⁵ position of C-ring and the addition of hydrophobic substituents on D-ring [15]. The phenolic derivatives were found to inhibit ABCG2 even more potently and selectively, and they markedly stimulated ATPase activity, by contrast with ketonic derivatives, or with Ko 143 and chromones that strongly inhibited ATPase activity [16]. In view of future use in animal models, the ADMET parameters of the most potent inhibitors were analyzed, to select those having the best profile.

Since 2007, in order to minimize animal testings, REACH regulation (regulation for Registration, Evaluation, Authorisation and restriction of CHemicals) promotes a “reduction, refinement or replacement” of animal use (3R) with alternative methods (annexe VII to XI) to evaluate toxicity of chemicals [17]. The principle of the “Three Rs” has been present early (1986) in the EU, since 1991 the ECVAM (European Centre for the Validation of Alternative Methods) is implicated in the scientific validation, the promotion of development and dissemination of new methods. The European Medicine Agency’s (EMA), in charge since 1995 of the scientific evaluation, supervision and safety monitoring of medicines in the EU, promotes the regulatory acceptance of 3R by the replacement of animal studies with *in vitro* models. On January 1, 2013 the EU Directive 2010/63/EU was followed in July 2013 by the creation of NETVAL (European Union Network of Laboratories for the Validation of Alternative Methods) to support the EURL ECVAM actions.

Alternative methods include *in silico* studies [17], with the advantage to be less expensive and time-consuming. These methods give a high throughput, could be optimized and require a lower synthesis of compounds. Therefore, poor pharmacokinetic profiles (ADME) can be detected early, avoiding costly late-stage failure in drug development.

The ECHA (European CHemical Agency) Guidance on information requirements and chemical safety assessment identified three main approaches for getting *in silico* non-testing data: (1) grouping approaches, including read-across and chemical category formation; (2) structure–activity relationship (SAR) and quantitative SAR, and (3) expert systems [18].

In our study, three *in silico* prediction tools were used to determine physicochemical and ADMET properties of a set of twelve indenoindoles, six ketonic and six phenolic derivatives (**Figure 1**) in order to select the best one(s) for future *in vivo* tests.

Claude Bernard Lyon 1, SFR Santé Lyon - Est CNRS UMS3453 - INSERM US7, 8 Avenue Rockefeller, F-69373, Lyon Cedex 8, France.

Tel: +33478777542

Citation: Guragossian N, Gozzi GJ, Fouillet B, et al. Candidate Molecule Selection Based on *In Silico* Predicted ADMET Properties of 12 Indenoindole Derivatives. Chem Inform. 2016, 2:2.

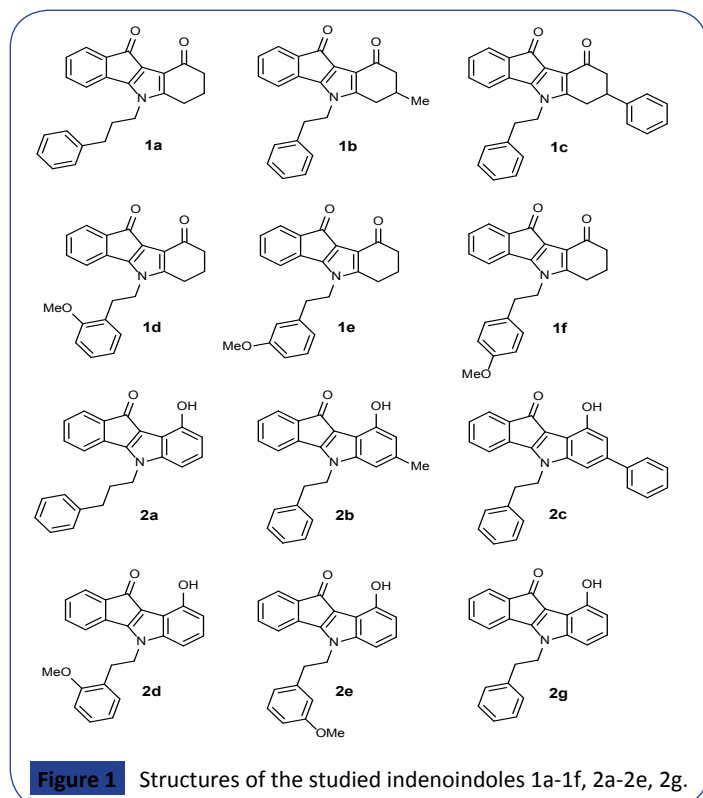
Materials and Methods

Physicochemical properties

Altogether 12 physicochemical parameters such as cLogP, molecular weight, TPSA (topological polar surface area) value, the number of hydrogen acceptor and donor and rotatable bonds have been calculated and are predictable by the Molinspiration software (v2013.09) [19]. cLogP parameter is defined as the sum of fragment-based contributions and correction factors. TPSA calculation is based on the summation of tabulated surface contributions of polar fragments [20]. These fragment contributions were determined by least-squares fitting to the single conformer 3D PSA for 34,810 drugs from the World Drug Index (correlation coefficient ≈ 0.99) [20].

ADME parameters

The ADME parameters (n=12) have been calculated by the ACD/Percepta software (v.14.0.0) [21]. The model used for the bioavailability module is based on differential equations describing solubility in the gastrointestinal tract, passive absorption in jejunum and elimination. The first pass effect is not considered in simulation. The database has 790 compounds compiled from reference pharmacokinetic tabulations and various articles. The prediction studies for (i) protein binding module, and (ii) P450 substrate and regioselectivity module (metabolism), are based on GALAS (Global, Adjusted Locally According to Similarity) modeling methodology [22] which consists of a global Partial Least Squares (PLS) QSAR model corrected by predicted values of similar compounds to the tested compound 1 or 2 (**Figure 1**) that are in the training set (70% of the whole data set of the software). These results are completed by a reliability index (RI) ranging from 0 to 1, where 0 is unreliable for the prediction and 1 is a fully reliable prediction. This RI value tends to zero in two cases: i) when the overall similarity between the tested compound 1 or 2 and the most similar compounds used for the correction of the global model is weak, because the predicted values of each of these compounds (obtained by bootstrapping the training set 100 times) weakly correlate with those of 1 or 2, ii) when an inconsistent variability between the predicted and the experimental values was observed among the most similar compounds to the tested compound 1 or 2. The validation sets were used to evaluate result accuracy by computing the Root Mean Square Error (RMSE) between the studied parameter



experimental value and its final predicted value. The results with a reliability index under 0.3 (cut-off value) should be discarded without any consideration to the result.

Predictive parameters for blood brain barrier permeability: The predictive models of Log PS and Log BB constants were built by using non-linear least squares regression validated by an internal validation set, and two other experimental external validation sets [23,24]. Physicochemical properties like lipophilicity (Log P), the number of hydrogen bond acceptors and donors, the ion form fraction at pH 6.5, and McGowan volume were calculated with the ACD/Labs Algorithm Builder 1.8 development platform, and used for the modeling. The combination of brain/plasma equilibration rate ($\text{Log}(PS * \text{fraction unbound, brain})$) with the partitioning of compounds at equilibrium (Log BB) classifies compounds as either active or inactive on the central nervous system; the model is validated using experimental data of more than 1500 compounds having a CNS activity [25].

Toxicity Parameters

The toxicity parameters (n=12) have been calculated by the ACD/Percepta [21] and the QSAR Toolbox [26] softwares.

ACD/Percepta software (v.14.0.0)

This software uses a human expert rules system [27-29] to predict mutagenicity, clastogenicity, carcinogenicity and reproductive toxicity. The acute toxicity module for the prediction of intravenous and oral LD50 uses the Galas modelling methodology [22]. The model predicting "health effects on particular organ or organ system" is based on RTECS and ESIS databases for more than 100,000 compounds. The skin irritation

model was constructed by using binomial PLS method from a set of pre-selected fragments and physicochemical parameters as descriptors, some specific variables are used to discriminate ionizable compounds and their salts. The experimental data have been collected from RTECS and ESIS databases, and diverse publications.

QSAR Toolbox software (v.3.4)

In vitro Ames test, in vivo micronucleus test and protein binding in chromosomal aberration test: The *in vitro* Ames test is related to the mutagenicity/carcinogenicity module of the software Toxtree (v.2.3) based on 30 structural alerts excluding non-genotoxic carcinogenicity alerts. The *in vivo* micronucleus profiler is based on the Toxmics rule base of the software Toxtree, providing 35 structural alerts known to induct the *in vivo* rodent micronucleus assay positive according to the structure of compounds and their mechanism of action. The chromosomal aberration test highlighted 33 structural alerts accounting for interaction of chemicals with proteins like topoisomerase II or cellular protein adducts.

Carcinogenicity: The profiler is an expanded and updated version of Toxtree software (v.2.3) module. A list of 55 structural alerts is considered for the carcinogenicity risks. From that list, 35 alerts are classified genotoxic, the others newly added are for most of them related to non-genotoxic carcinogenicity.

Results and Discussion

Physicochemical properties

Calculations showed that all the twelve compounds displayed a topological polar surface area (TPSA) lower than 60 \AA^2 , and less than ten rotatable bonds, indicating that they could be good candidates upon oral administration (**Table 1**) [30]. All selected molecules were also predicted to exhibit a moderate brain uptake on the basis of TPSA values under 60 \AA^2 [31], and might then be active in the central nervous system.

The physicochemical parameters of Lipinski's rule of five indicated that a good absorption or permeation is more likely when molecules have a calculated LogP (cLogP) below 5, a molecular weight lower than 500, less than 10 H-bond acceptors and 5 H-bond donors [32]. All evaluated compounds indeed followed that rule (**Table 1**), except for the cLogP parameter, where the values were higher for one ketonic compound, 1c, and for the six phenolic compounds 2a-e and 2g. In contrast, compounds 1c, 2c presented a much higher cLogP value (6.18 and 6.90, respectively) showing a less favorable permeation profile. Nevertheless, the compounds having cLogP values up to 5.6 might have drug-likeness properties [33].

ADME parameters

Absorption: The results of oral bioavailability (**Table 2**) were in the range of 39-78%, which could be correlated to a good absorption rate after passive transcellular transport [34], except for compound 2c (with a predicted low value of 10%).

Distribution: The twelve compounds may have high affinity for plasmatic proteins, with values close to 99% binding (**Table 2**);

however, the results are in the model applicability domain for only ketonic compounds and one phenolic compound 2c, with a reliability index (RI) over 0.3 (from 0.31 to 0.43). All the compounds displayed Log[brain]/[blood] (Log BB) values between -0.76 and 0.14, and Log PS values from -1.1 to -1.3, suggesting a moderate distribution across the blood-brain barrier (**Table 2**) [35].

Metabolism: The compounds had a high probability (>0.90), with a good reliability (RI from 0.31 to 0.41) for the six phenolic derivatives 2a-2e, 2g and the two ketonic ones 1a, 1b of being metabolized by the CYP3A4 enzyme (**Table 2**). Such a modification could mainly consist of specific hydroxylation at para position of the phenethyl group for compounds 2b-2d and 2g.

Elimination: Ketonic and phenolic compounds had a similar elimination-rate constant (k_e), from 0.00082 to 0.0011; in contrast, compound 2c displayed a 2-fold lower k_e value, indicating a higher retention time inside the body (**Table 2**).

Toxicity parameters

Mutagenicity: The knowledge-based expert system highlighted the presence of a planar polycyclic system [29] among all

compounds (ketonic and phenolic derivatives) predicting that they all could be DNA-intercalating agents; moreover, some of unsubstituted aromatic regions may be DNA reactive after epoxidation by P450 cytochrome enzymes as mentioned in ACD percepta by "positive" in **Table 3**.

According to the QSAR Toolbox software, the mutagenicity risks (Ames test, without metabolic activation) were negative for all derivatives (**Table 3**). The software can also detect, with a SAR approach, alerts containing electrophilic centers or metabolically activated, attributed to one of six mechanistic domains related to mutagenicity or carcinogenicity [36]. A risk of a specific metabolism by P450 cytochrome enzymes for the ketonic derivatives was highlighted, leading to reactive iminium species and implicating the formation of DNA adducts via a SN1 mechanism [37], but this metabolic pathway was not confirmed by the ACD/Percepta software. On the other hand, an epoxidation mediated by P450 cytochrome enzymes on phenethyl benzenic group was predicted for the three ketonic compounds 1a-1c and the four phenolic ones 2a-2c and 2g. The compounds could then

Table 1 Physico-chemical parameters of the 12 indenoindoles predicted by the Molinspiration software.

Cpds	Studied parameters					
	cLogP ^a	Molecular weight (g/mol)	Acceptor H-bonds	Donor H-bonds	TPSA (Å ²) ^b	Rotatable bonds ^c
1a	4.90	355.44	3	0	39.07	4
1b	4.86	355.44	3	0	39.07	3
1c	6.18	417.51	3	0	39.08	4
1d	4.39	371.44	4	0	48.31	4
1e	4.42	371.44	4	0	48.31	4
1f	4.44	371.44	4	0	48.31	4
2a	5.67	353.42	3	1	42.23	4
2b	5.55	353.42	3	1	42.23	3
2c	6.90	415.49	3	1	42.23	4
2d	5.16	369.42	4	1	51.47	4
2e	5.18	369.42	4	1	51.46	4
2g	5.15	339.39	3	1	42.23	3

^a cLogP: calculated logarithm of partition coefficient of a compound between octanol and water. ^b TPSA: topological polar surface area (Å²). ^c Rotatable bonds: single non-ring bond, bounded to nonterminal heavy atom (excepted for C-N bonds).

Table 2 ADME parameters of the 12 indenoindoles predicted by the ACD/Percepta software.

Cpds	ADME parameters					
	Bioavailability (%)	% PPB ^a	Log BB ^b	Log PS ^c	Metabolism by CYP3A4 enzyme ^d	Elimination rate constant (min ⁻¹)
1a	61	99.26	0.08	-1.1	0.96	0.001
1b	70	99.34	0.02	-1.1	0.98	0.001
1c	39	99.90	-0.68	-1.1	0.99	0.00094
1d	78	99.06	-0.02	-1.1	0.96	0.0011
1e	78	99.06	-0.02	-1.1	0.98	0.0011
1f	78	99.06	-0.02	-1.1	0.96	0.0011
2a	58	99.22	0.13	-1.1	0.95	0.00082
2b	52	99.18	0.13	-1.1	0.95	0.00083
2c	10	99.91	-0.76	-1.3	0.98	0.00044
2d	62	99.12	0.10	-1.1	0.97	0.00086
2e	62	99.12	0.10	-1.1	0.96	0.00086
2g	61	99.14	0.14	-1.1	0.94	0.00086

^a % PPB = (1-fu)x100% (percentage of compounds binding to plasmatic proteins where fu represents the unbound fraction of compounds).

^b Log BB: logarithm of steady-state distribution ratio of a compound between brain tissue and plasma. ^c Log PS: logarithm of permeability surface area coefficient ($PS = -F * (1 - e^{-K_{in}/F})$, with F: blood flow rate in cerebral micro capillaries and K_{in} : influx rate constant). ^d Probability of compounds to be metabolized by the CYP3A4 enzyme.

Table 3 Toxicity parameters of the 12 indenoindoles predicted by the ACD/Percepta and QSAR Toolbox softwares.

Cpd	Toxicity parameters															
	Mutagenicity			Clastogenicity			Carcinogenicity ^d			Binding to ER-alpha ^e (-3<Log RBA ≤ 0)						
	ACD/Percepta ^a	QSAR Toolbox ^a	QSAR Toolbox	ACD/Percepta ^c	QSAR Toolbox	Micro-nucleus test	Protein binding	ACD/Percepta	QSAR Toolbox	Hepato-toxicity ^f	Lung toxicity ^g	Blood toxicity ^h	LD ₅₀ IV	LD ₅₀ per os	Skin toxicity	
1a	positive	negative	yes [quinone, iminium ion] ^b	inconclusive	positive	positive	negative	inconclusive	negative	yes	0.35	0.50	0.09	36	980	0.28
1b	positive	negative	yes [quinone, iminium ion]	inconclusive	positive	positive	negative	inconclusive	negative	yes	0.35	0.43	0.08	33	850	0.28
1c	positive	negative	yes [quinone, iminium ion]	inconclusive	positive	positive	negative	inconclusive	negative	yes	0.33	0.96	0.56	27	1100	0.26
1d	positive	negative	yes [iminium ion]	inconclusive	positive	positive	negative	inconclusive	negative	yes	0.34	0.93	0.03	36	880	0.28
1e	positive	negative	yes [iminium ion]	inconclusive	positive	positive	negative	inconclusive	negative	yes	0.34	0.93	0.05	36	880	0.28
1f	positive	negative	yes [iminium ion]	inconclusive	positive	positive	negative	inconclusive	negative	yes	0.34	0.93	0.05	36	880	0.28
2a	positive	negative	yes [quinone]	inconclusive	positive	positive	negative	inconclusive	negative	yes	0.20	0.42	0.21	28	540	0.28
2b	positive	negative	yes [quinone]	inconclusive	positive	positive	positive	inconclusive	negative	yes	0.26	0.45	0.19	28	560	0.28
2c	positive	negative	yes [quinone]	inconclusive	positive	positive	positive	inconclusive	negative	yes	0.19	0.96	0.31	21	480	0.26
2d	positive	negative	no alert	inconclusive	positive	positive	negative	inconclusive	negative	yes	0.15	0.40	0.50	30	560	0.28
2e	positive	negative	no alert	inconclusive	positive	positive	negative	inconclusive	negative	yes	0.23	0.40	0.13	30	560	0.28
2g	positive	negative	yes [quinone]	inconclusive	positive	positive	negative	inconclusive	negative	yes	0.21	0.47	0.18	31	580	0.29

^aMutagenicity was determined by Ames test. ^bStructural alert for DNA binding risks after metabolic activation. ^cClastogenicity was evaluated by both chromosome aberrations and micronucleus *in vivo* composite tests. ^dThe carcinogenicity test is concerning mice and other rodents' species. ^eLog RBA = Log (IC₅₀ (17β-estradiol) × 100/IC₅₀ (compound)) where IC₅₀ is the concentration where a ligand displaces half of specifically bound 17β-estradiol from ER-alpha. ^fRBA: ^gProbability of compounds having chronic, sub-chronic or acute toxic effects on liver, lung and blood, respectively.

be converted to reactive quinones, and a nucleophilic attack by the nitrogen of DNA could be effective through a Michael addition mechanism, leading to DNA adducts [38]. The compounds 1c and 2c could have a higher probability of being metabolized with an additional phenyl group at position C7. However, derivatives with a methoxy substituent on the phenethyl group (compounds 2d, 2e) were not predicted to be metabolized, maybe due to steric hindrance on specific sites, excluding them from any potential mutagenic risk.

The DNA-intercalating risk highlighted by softwares for our compounds is a feature of heteropolycyclic aromatic hydrocarbons because of their planar polycyclic system that could make stacking interaction with nucleotides, leading to mutagenicity. The crystallographic structure of phenolic indenoindole derivatives showed that their structure was nearly planar, with a mean out-of-plane deviation of 0.027 Å, determined by the derivation of the C12 and C13 atoms from the plane defined by the heterotetracyclic system [16], contrary to the ketonic derivatives with values of 0.1935 Å and 0.4429 Å, respectively [15]. In the literature, some indenoindole analogues of phenolic sub-series were found to be DNA-intercalating agents [39]. Nevertheless, an aromatic polycyclic structure alone could be insufficient to get that property [40], and this capacity seems to be mostly dependent on the type of substitutions. For some compounds, the presence of a cationic amino-alkylated structural alert might enhance the affinity of compounds for the negatively-charged phosphodiester backbone of DNA, and then facilitate their intercalating property by increasing DNA affinity through electrostatic interactions [41]. The addition of a N-dialkyl group and its absence for triphenylethylene derivatives implicates a lack of DNA intercalation [42], and the incorporation of an extended alkylamine group is sometimes used to functionalize anti-cancer therapeutic intercalating agents (e.g., mitoxantrone), or inhibitors of transcription factors [43].

In our case, we noticed that the twelve compounds have a pyrrole moiety and a predicted pKa value, by the ACD/Percepta software, around 9 (data not shown). This indicates that molecules could be positively charged at physiological pH. However, the conjugated system of the pyrrole group makes its protonation difficult, and implicates that compounds may have lower intercalating properties.

Clastogenicity: No conclusions could be drawn for *in vivo* clastogenicity risks with the ACD/Percepta software. For all compounds, the QSAR Toolbox software detected a structural alert representing two atoms connected to hydrogen-bond acceptors. Therefore, the derivatives could interact with DNA via non-covalent binding, acting as DNA groove binders [44] with the formation of a drug/DNA complex, leading to higher risks of replicative errors.

Another characteristic conducting to clastogenicity risks is the capacity of molecules to interact with proteins like topoisomerase II during DNA replication. This property may occur when compounds are able to form a stable complex between DNA and topoisomerase II, leading to DNA strand breaks without religation, acting as topoisomerase II poisons. Many intercalating drugs

like anthracyclines (e.g., doxorubicin) [45], or non-intercalating compounds such as epipodophyllotoxin etoposide [46] can form topo II-DNA covalent complex conferring clastogenic properties.

QSAR Toolbox highlighted a risk for the two compounds 2b, 2c to interact with topoisomerase II (**Table 3**). This process may implicate a perhydroxylation reaction, resulting from the attack of molecular oxygen in ortho- or para-phenolic position, and the formation of a free peroxy radical stabilized by the methyl (2b) or phenyl (2c) substituent and then converted into the corresponding perhydroxylated derivative. This step is followed by the formation of the ortho- and para-benzoquinones formed as a result of an intramolecular dehydration. Alkylating cellular proteins such as sulfur nucleophiles and other nucleophilic sites in proteins could then attack the quinoid structures via a Michael-type addition reaction leading to the possible clastogenic effects.

Carcinogenicity: No conclusion was possible for carcinogenicity risks on mice and other rodents with the ACD/Percepta software, while QSAR Toolbox detected no structural alerts for that risk (**Table 3**).

Other toxicity parameters: The twelve compounds were predicted to have moderate-to-weak binding to the ER-alpha (estrogen receptor) [47] with log RBA values between -3 and 0, suggesting a low toxicity on reproduction (**Table 3**). The interaction between indenoindole derivatives and the estrogen receptor could be related to the presence of (i) the phenolic ring and (ii) additional ring structures [48], and then explain the relative toxicity.

No significant toxicity was found on the liver or the blood system (except for 1c). A potential damage on lungs could be predicted for compounds 1c-f and 2c, the probability being lower or equal to 0.50 for 1a, 1b, 2a, 2b, 2d, 2e and 2g. The twelve compounds have LD₅₀ values from 480 to 1100 mg/kg, classifying them in the category 4 of acute toxicity and implicating a slight toxicity for all of them. No skin toxicity was observed for any compound (**Table 3**).

Conclusion

For the ACD/Percepta software, the developers defined three threshold RI values which subdivided the reliability of predictions into four categories: unreliable (lower than 0.3), borderline (0.3), moderate (0.5) and reliable (0.75). In the case of QSAR LD₅₀ model for RI values higher than 0.5 or 0.75, the root mean square error (RMSE) fell below 0.5 log units, which is considered to be within the inter-laboratories measurement error. According to the software, falling in this range could give a good estimation of acute toxicity results for compounds, and remained close to *in vitro* testing in the early stages of drug development.

With our set of compounds, no predictions had a reliability index over 0.75, and only some of them related to LD₅₀ values were over 0.5. A low RI value implicates a lack of compounds similarity in the training set or an inconstancy of baseline QSAR model predictions with experimental data of five structurally-related analogues with the test compound. This highlights the limitations of QSAR models based on a structurally-predefined training set, where a change in property or mechanism of action of analogues

might be sometimes linked to slight structural changes. To address such a problem, QSAR models of the ACD/Percepta software can be trained by including new experimental data for similar compounds to expand the chemical space and the applicability domain [49], without any statistical reparameterization of the model.

Expert system results need to be analyzed carefully and, as previously seen about DNA intercalation risks, it is necessary to have many structural alerts and to consider the results as a weight of evidence to be able to conclude to a significant effect.

With the present studies, we could exclude compound 2c for relatively low oral bioavailability, 1c for blood toxicity and 1c-f and 2c for pulmonary toxicity. The two phenolic derivatives 2d, 2e should not be metabolically activated by CYP enzymes according to the QSAR Toolbox software leading to a lower mutagenic risk.

Compounds 2b, 2c might be clastogenic on the basis of a hydrogen bond acceptor structural alert with a possible interaction with topoisomerase II, but no conclusion could be given about carcinogenicity risks.

In conclusion, compounds 1a, 1b, 2a, 2d, 2e, 2g may have an acceptable pharmacokinetic and toxicity profile; nevertheless, some preclinical *in vitro* experiments could help us to confirm mutagenicity and clastogenicity risks (Ames test with metabolic activation, chromosome aberration tests), or evaluation of organ toxicity with *in vitro* specific cytotoxicity tests.

Acknowledgements

The authors would like to thank the ACD/Labs Company for advices concerning the ACD/Percepta software.



References

- 1 Juliano RL, Ling V (1976) A surface glycoprotein modulating drug in chinese hamster ovary cell mutants. *Biochim Biophys Acta* 455: 152-162.
- 2 Cole SP, Bhardwaj G, Gerlach JH, Mackie JE, Grant CE, et al. (1992) Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science* 258: 1650-1654.
- 3 Allikmets R, Schriml LM, Hutchinson A, Romano-Spica V, Dean M (1998) A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance. *Cancer Res* 58: 5337-5339.
- 4 Doyle LA, Yang W, Abruzzo LV, Krogmann T, Gao Y, et al. (1998) A multidrug resistance transporter from human MCF-7 breast cancer cells. *Proc Natl Acad Sci USA* 95: 15665-15670.
- 5 Miyake K, Mickley L, Litman T, Zhan Z, Robey R, et al. (1999) Molecular cloning of cDNAs which are highly overexpressed in mitoxantrone-resistant cells: demonstration of homology to ABC transport genes. *Cancer Res* 59: 8-13.
- 6 Diestra JE, Scheffer GL, Catala I, Maliepaard M, Schellens JH, et al. (2002) Frequent expression of the multi-drug resistance-associated protein BCRP/MXR/ABCP/ABCG2 in human tumours detected by the BXP-21 monoclonal antibody in paraffin-embedded material. *J Pathol* 198: 213-219.
- 7 Rabindran SK, He H, Singh M, Brown E, Collins KI, et al. (1998) Reversal of a novel multidrug resistance mechanism in human colon carcinoma cells by fumitremorgin C. *Cancer Res* 58: 5850-5858.
- 8 Allen JD, van Loevezijn A, Lakhai JM, van der Valk M, van Tellingen O, et al. (2002) Potent and specific inhibition of the breast cancer resistance protein multidrug transporter *in vitro* and in mouse intestine by a novel analogue of fumitremorgin C. *Mol Cancer Ther* 1: 417-425.
- 9 Nicolle E, Boumendjel A, Macalou S, Genoux E, Ahmed-Belkacem A, et al. (2009) QSAR analysis and molecular modeling of ABCG2-specific inhibitors. *Adv Drug Deliver Rev* 61: 34-46.
- 10 Boumendjel A, Macalou S, Valdameri G, Pozza A, Gauthier C, et al. (2011) Targeting the multidrug ABCG2 transporter with flavonoid inhibitors: *in vitro* optimization and *in vivo* validation. *Curr Med Chem* 18: 3387-3401.
- 11 Lecerf-Schmidt F, Peres B, Valdameri G, Gauthier C, Winter E, et al. (2013) ABCG2: recent discovery of potent and highly selective inhibitors. *Future Med Chem* 5: 1037-1045.
- 12 Winter E, Locatelli C, Di Pietro A, Creczynski-Pasa TB (2015) Recent trends of the chalcones potentialities as antiproliferative and antiresistance agents. *Anticancer Agents Med Chem* 15: 592-604.
- 13 Arnaud O, Boumendjel A, Geze A, Honorat M, Mattera E, et al. (2011) The acridone derivative MBLI-87 reverses breast cancer resistance protein-mediated resistance to irinotecan in xenografts. *Eur J Cancer* 47: 640-648.
- 14 Honorat M, Guitton J, Gauthier C, Bouard C, Lecerf-Schmidt F, et al. (2014) MBL-II-141, a chromone derivative, enhances irinotecan anticancer efficiency in ABCG2-positive xenografts. *Oncotarget* 5: 11957-11970.
- 15 Jabor Gozzi G, Bouaziz Z, Winter E, Daflon-Yunes N, Aichele D, et al. (2015) Converting potent indeno[1,2-b]indole inhibitors of protein kinase CK2 into selective inhibitors of the breast cancer resistance protein ABCG2. *J Med Chem* 58: 265-277.
- 16 Gozzi GJ, Bouaziz Z, Winter E, Daflon-Yunes N, Honorat M, et al. (2015) Phenolic indeno[1,2-b]indoles as ABCG2-selective potent and non-toxic inhibitors stimulating basal ATPase activity. *Drug Des Devel Ther* 9: 3481-3495.
- 17 Regulation (EC) No 1907/2006. REACH - Registration, evaluation, authorisation and restriction of chemicals. http://ec.europa.eu/enterprise/sectors/chemicals/reach/index_en.htm
- 18 ECHA (2008) Guidance on information requirements and chemical safety assessment chapter R.6: QSARs and grouping of chemicals. Available at: <http://guidance.echa.europa.eu/docs/>
- 19 For more information on Molinspiration Software, use the corresponding link: <http://www.molinspiration.com/cgi-bin/properties> Accession date: December 12, 2014.
- 20 Ertl P, Rohde B, Selzer P (2000) Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J Med Chem* 43: 3714-3717.
- 21 ACD/Percepta (2014) Version 14.0.0. Advanced Chemistry Development Inc., Toronto, ON, Canada, www.acdlabs.com
- 22 Sazonovas A, Japertas P, Didziapetris R (2010) Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD 50). *SAR QSAR Environ Res* 21: 127-148.
- 23 Kalvass JC, Maurer TS, Pollack GM (2007) Use of plasma and brain unbound fractions to assess the extent of brain distribution of 34 drugs: comparison of unbound concentration ratios to *in vivo* p-glycoprotein efflux ratios. *Drug Metab Dispos* 35: 660-666.
- 24 Summerfield SG, Lucas AJ, Porter RA, Jeffrey P, Gunn RN, et al. (2008) Towards an improved prediction of human *in vivo* brain penetration. *Xenobiotica* 38: 1518-1535.
- 25 Adenot M, Lahana R (2004) Blood-brain barrier permeation models: discriminating between potential CNS and non-CNS drugs including P-glycoprotein substrates. *J Chem Inf Comput Sci* 44: 239-248.
- 26 For more information on the QSAR Toolbox, use the corresponding link: <http://www.oecd.org/env/ehs/risk-assessment/theoecdqsartoolbox.htm> Accession date: October 05, 2015.
- 27 Benigni R, Bossa C (2006) Structural alerts of mutagens and carcinogens. *Curr Comput Aided Drug Des* 2: 169-176.
- 28 Kazius J, McGuire R, Bursi R (2005) Derivation and validation of toxicophores for mutagenicity prediction. *J Med Chem* 48: 312-320.
- 29 Kazius J, Nijssen S, Kok J, Bäck T, Ijzerman AP (2006) Substructure mining using elaborate chemical representation. *J Chem Inf Model* 46: 597-605.
- 30 Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, et al. (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 45: 2615-2623.
- 31 Kelder J, Grootenhuys PDJ, Bayada DM, Delbressine LP, Ploemen JP (1999) Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm Res* 16: 1514-1519.
- 32 Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46: 3-26.
- 33 Ghose AK, Viswanadhan VN, Wendoloski JJ (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry

- libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J Comb Chem* 1: 55-68.
- 34 Tingjun H, Junmei W, Wei Z, Xiaojie X (2007) ADME Evaluation in drug discovery. 7. Prediction of oral absorption by correlation and classification. *J Chem Inf Model* 47: 208-218.
- 35 Abraham MH, Takács-Novák K, Mitchell RC (1997) On the partition of ampholytes: application to blood-brain distribution. *J Pharm Sci* 86: 310-315.
- 36 Aptula AO, Roberts DW (2006) Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. *Chem Res Toxicol* 19: 1097-1105.
- 37 Kalgutkar AS, Gardner I, Scott Obach R, Shaffer CL, Callegari E, et al. (2005) A comprehensive listing of bioactivation pathways of organic functional groups. *Curr Drug Metab* 3: 161-225.
- 38 Volk DE, Thiviyanathan V, Rice JS, Luxon BA, Shah JH, et al. (2003) Solution structure of a cis-opened (10R)-N6 deoxyadenosine adduct of (9S,10R)-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene in a DNA duplex. *Biochemistry* 42: 1410-1420.
- 39 Bal C, Baldeyrou B, Moz F, Lansiaux A, Colson P, et al. (2004) Novel antitumor indenoindole derivatives targeting DNA and topoisomerase II. *Biochem Pharmacol* 68: 1911-1922.
- 40 Kashyap M, Kandekar S, Baviskar AT, Das D, Preet R, et al. (2013) Indenoindolone derivatives as topoisomerase II-inhibiting anticancer agents. *Bioorg Med Chem Lett* 23: 934-938.
- 41 Bruice TC, Mei HY, He GX, Lopez V (1992) Rational design of substituted tripyrrole peptides that complex with DNA by both selective minor-groove binding and electrostatic interaction with the phosphate backbone. *Proc Natl Acad Sci USA* 89: 1700-1704.
- 42 Snyder RD, Brown JE (2002) Evidence for and role of the dimethylamino group in tamoxifen DNA intercalation in intact Chinese hamster V79 cells. *Drug Chem Toxicol* 25: 473-479.
- 43 Satz AL, Bruice TC (2002) Recognition in the minor groove of double-stranded DNA by microgonotropens. *Acc Chem Res* 35: 86-95.
- 44 Snyder RD, Ewing D, Hendry LB (2006) DNA intercalative potential of marketed drugs testing positive in *in vitro* cytogenetics assays. *Mutat Res* 609: 47-59.
- 45 Ferguson LR (1998) Inhibitors of topoisomerase II enzymes: a unique group of environmental mutagens and carcinogens. *Mutat Res* 400: 271-278.
- 46 Ashby J, Tinwell H, Glover P, Poorman-Allen P, Krehl R, et al. (1994) Potent clastogenicity of the human carcinogen etoposide to the mouse bone marrow and mouse lymphoma L5178Y cells: comparison to Salmonella responses. *Environ Mol Mutagen* 24: 51-60.
- 47 Blair RM, Fang H, Branham WS, Hass BS, Dial SL, et al. (2000) The estrogen receptor relative binding affinities of 188 natural and xenochemicals: structural diversity of ligands. *Toxicol Sci* 54: 138-153.
- 48 Fang H, Tong W, Shi LM, Blair R, Perkins R, et al. (2001) Structure-activity relationships for a large diverse set of natural, synthetic, and environmental estrogens. *Chem Res Toxicol* 14: 280-294.
- 49 Didziapetris R, Dapkunas J, Sazonovas A, Japertas P (2010) Trainable structure-activity relationship model for virtual screening of CYP 3A4 inhibition. *J Comput Aided Mol Des* 24: 891-906.