HHS Public Access

Predicting chemically-induced skin reactions. Part II: QSAR models of skin permeability and the relationships between skin permeability and skin sensitization

Vinicius M. Alves[a,b], **Eugene Muratov**[b,c], **Denis Fourches**[b], **Judy Strickland**[d], **Nicole Kleinstreuer**[d], **Carolina H. Andrade**[a], and **Alexander Tropsha**[b,*]

[a]Laboratory of Molecular Modeling and Design, Faculty of Pharmacy, Federal University of Goiás, Goiânia, GO, 74605-220, Brazil

[b]Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA

[c]Laboratory of Theoretical Chemistry, A.V. Bogatsky Physical-Chemical Institute NAS of Ukraine, Odessa, 65080, Ukraine

[d]ILS/Contractor supporting the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), P.O. Box 13501, Research Triangle Park, NC, 27709, USA

Abstract

Skin permeability is widely considered to be mechanistically implicated in chemically-induced skin sensitization. Although many chemicals have been identified as skin sensitizers, there have been very few reports analyzing the relationships between molecular structure and skin permeability of sensitizers and non-sensitizers. The goals of this study were to: (*i*) compile, curate, and integrate the largest publicly available dataset of chemicals studied for their skin permeability; (*ii*) develop and rigorously validate QSAR models to predict skin permeability; and (*iii*) explore the complex relationships between skin sensitization and skin permeability. Based on the largest publicly available dataset compiled in this study, we found no overall correlation between skin permeability and skin sensitization. In addition, cross-species correlation coefficient between human and rodent permeability data was found to be as low as $R^2=0.44$. Human skin permeability models based on the random forest method have been developed and validated using OECD-compliant QSAR modeling workflow. Their external accuracy was high ($Q^2_{ext} = 0.73$ for 63% of external compounds inside the applicability domain). The extended analysis using both experimentally-measured and QSAR-imputed data still confirmed the absence of any overall

[*]**Corresponding Author**. Address for correspondence: 100K Beard Hall, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA; Telephone: (919) 966-2955; FAX: (919) 966-0204; alex_tropsha@unc.edu.

CONFLICT OF INTERESTS
The authors declare no actual or potential conflict of interests.

concordance between skin permeability and skin sensitization. This observation suggests that chemical modifications that affect skin permeability should not be presumed a priori to modulate the sensitization potential of chemicals. The models reported herein as well as those developed in the companion paper on skin sensitization suggest that it may be possible to rationally design compounds with the desired high skin permeability but low sensitization potential.

Keywords

Skin sensitization; skin permeability; QSAR; virtual screening; skin toxicants

INTRODUCTION

Skin sensitization is a complex adverse toxicological endpoint that is influenced by several biological parameters, such as protein binding, dendritic cell activation, individual variation, and time-dose exposure (Jaworska et al., 2011; Johansen et al., 2011). Skin permeability is also often considered as a potential parameter affecting chemicals' sensitization potential (MacKay et al., 2013). It relates to the ability of a molecule to pass through the skin, a characteristic which is primarily influenced by the physicochemical properties of the chemical as well as the physicochemical and biological properties of the membrane (Xia, 2011).

Despite the high importance of skin permeability for consumer product efficacy and its supposed influence on potential toxicities such as skin sensitization, the amount of experimental data available in the public domain is surprisingly limited. In the early 90's, a compilation of data points for skin permeability gathered from several sources was published (Flynn, 1990). Subsequent studies added some complementary data allowing researchers to develop Quantitative Structure-Activity Relationship (QSAR) models for predicting skin permeability. However, as shown in Table S1, most of the published studies reporting on skin permeability modeling (Abraham et al., 1999; Barratt, 1995; Berge, 2009; Chen et al., 2010, 2007; Cronin et al., 1999; Hostýnek and Magee, 1997; Lien and Gao, 1995; Magnusson et al., 2004; Moss and Cronin, 2002; Moss et al., 2011; Patel et al., 2002; Potts and Guy, 1995, 1992) have not included certain critical elements of QSAR model development and validation protocol, such as the definition of the applicability domain (AD) or proof of passing the Y-randomization test, which constitute best practices of QSAR modeling (OECD, 2004; Tropsha, 2010). Recently, several QSAR studies were benchmarked on a series of 11 compounds and all of them failed to predict skin permeation quantitatively; they were only able to rank permeants (Brown et al., 2012). Another recent study showed that most of the available QSAR models underestimate the skin permeability of hydrophilic solutes (Chen et al., 2013). Recent studies (Chauhan and Shakya, 2010) attempted to overcome the related problems with complex chemical mixtures and built a QSAR model based on several mixtures of 36 chemicals with porcine skin data. The latter model followed the best practices of QSAR modeling (Tropsha, 2010); however, we have identified 21 duplicative structures in that dataset (see Table S3) indicating a potential bias of the model and a likely over-estimation of its true performance.

Despite the underlying importance of skin permeability and its identification as a necessary step in the OECD Adverse Outcome Pathway (AOP) for skin sensitization (Karlberg et al., 2008; OECD, 2012), we could not find any study among those compiled in Tables S1 and S2, where both endpoints were analyzed concurrently and in the context of their possible inter-dependency. The prevalence of dermal exposure to diverse chemicals in consumer products and in the environment, the importance of permeability for skin sensitization, and the lack of reliable models to predict these endpoints for new chemicals have motivated us to initiate a tandem study on collecting and analyzing both skin permeability and skin sensitization data. In the companion paper (Alves et al., 2014), we have reported on new QSAR models of skin sensitization. In this study, we have compiled, curated, and integrated skin permeability coefficient (Kp) data extracted from various literature sources. Using this unique data collection, we have developed and rigorously validated QSAR models for skin permeability, and explored the relationships between the skin sensitization potential and the chemical permeability coefficient. The QSAR models developed in this and the accompanying study (Alves et al., 2014) are publicly available and can be used for evaluating chemically induced skin effects *in silico* as part of both research and development projects as well as in support of regulatory decisions on consumer products.

MATERIALS AND METHODS

Datasets

Skin sensitization datasets (datasets A and B)—In the Part I of this study (Alves et al., 2014) we described two skin sensitization datasets. Briefly, one of them (dataset A) was retrieved from the ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) report on the murine reduced local lymph node assay (ICCVAM 2009). The modeling set (Dataset A) consisted of 254 compounds (127 sensitizers and 127 non-sensitizers) and the external validation set (dataset B) consisted of 133 sensitizers from the ICCVAM report (ICCVAM 2009) and 18 additional compounds taken from the study of Jaworska et al. (2011). This collection of data was used to explore the intrinsic relationship between skin sensitization and skin permeability (human data from dataset D; see below) for a subset of 20 compounds from the same dataset for which both skin sensitization and skin permeability data were known.

Human skin permeability dataset (dataset D)—*In vitro* human skin permeability coefficients were retrieved from the literature (Chauhan and Shakya, 2010) including 211 records expressed in logKp (cm.h^{-1}); this dataset contained the well-known and frequently studied Flynn dataset (Flynn, 1990). 17 duplicates and two sets of triplicates were identified and curated leaving unique compounds only. Three additional compounds and water were also removed for the following reasons: both styrene (logKp = −0.19) and ethyl benzene (logKp = 0.08) were identified as activity outliers (*i.e.*, their logKp values were far outside the activity range of the other compounds in the dataset between −5.52 and −0.69) whereas digitoxin (natural product) was a structural outlier. The remaining 186 compounds (dataset D) were retained for modeling.

Rodent dataset (dataset E)—The set of *in vitro* rodent skin permeability data consisting of 103 chemical compounds was retrieved from the literature (Moss et al., 2011). After curation, 96 compounds (dataset E) were kept for modeling. The following five activity outliers were removed from the dataset E: bisphenol A diglycidyl ether (–5.26), decabromodiphenyl oxide (–5.15), 4-N butylamine (–0.64), bufexamac (–0.57), and triclosan (0.13). The overall range of logKp for the final dataset varied from –4.85 to –0.94.

Data curation

Chemical structures were retrieved either from PubChem (https:// pubchem.ncbi.nlm.nih.gov/, accessed in March 2012) or ChemSpider (http:// www.chemspider.com/, accessed in March 2012) databases using chemical names. Chemicals were removed if their structures could not be found. Each dataset was carefully curated according to previously established guidelines (Fourches et al., 2010). Briefly, counterions were removed, whereas specific chemotypes such as aromatic and nitro groups were normalized using the ChemAxon Standardizer (v.5.3, ChemAxon, Budapest, Hungary, http://www.chemaxon.com). The presence of duplicates, *i.e.*, identical compounds reported more than once in the same dataset, is known to lead to over-optimistic estimations of the predictivity for developed QSAR models. However, the analysis of such records also gives an estimate of the dataset quality: if activity data for the same compound are consistent, the overall data quality is high; if there is a large deviation in experimental values between different records of the same compound, the quality is, obviously, low. Thus, after the structural standardization, the duplicates were identified using ISIDA Duplicates (Varnek et al., 2008) and HiT QSAR (Kuz'min et al., 2008) software and carefully analyzed. If the experimental properties associated with two duplicative structures were identical, or highly similar, then one compound was chosen at random and deleted. However, if their experimental properties were significantly different, we deleted both records from the dataset. During the curation of human skin permeability dataset, we found 17 pairs of duplicate structures and two sets of triplicates as shown in Table S3. The permeability values for duplicative records were almost identical (logKp variation ~ 0.01 LU) except those for butanoic acid that had a variation higher than ca. 0.4 LU. Thus, the LogKp value for all 17 compounds was averaged and the duplicative records were removed so that only one permeability coefficient was used for a chemical compound.

Cheminformatics Approaches

Hierarchical Cluster Analysis—The clustering of a chemical dataset consists of merging compounds into distinct clusters of chemically similar molecules [see (Downs and Barnard, 2003; Mercier, 2003) for a review of the most popular clustering approaches used in computational chemistry]. In this study, we have employed the Sequential Agglomerative Hierarchical Non-overlapping (SAHN) method implemented in the ISIDA/Cluster program (http://infochim.u-strasbg.fr) (Varnek et al., 2007). Briefly, each compound represents one cluster at the start. Then, the *m* compounds are merged iteratively into clusters using their pairwise Euclidean distances stored in a squared ($m * m$) symmetric distance matrix. The two closest objects (molecules or clusters) are iteratively identified and merged to form a new cluster, the distance matrix being updated with the re-computed distances separating the newly formed cluster and the others, according to the user-specified type of linkage

(complete linkage in this study). The process is repeated until one cluster remains. The parent-child relationships between clusters result in a hierarchical data representation, or *dendrogram*. In particular, we used ISIDA/Cluster to obtain the heat map (see Results section) of the proximity matrix.

Dragon Descriptors—The following types of descriptors were generated using Dragon software (v.5.5, Talete SRL, Milan, Italy): 0D constitutional (atom and group counts), 1D functional groups, 1D atom-centered fragments, 2D topological descriptors, 2D walk and path counts, 2D autocorrelations, 2D connectivity indices, 2D information indices, 2D topological charge indices, 2D eigenvalue-based indices, 2D topological descriptors, 2D edge-adjacency indices, 2D Burden eigenvalues, 2D binary fingerprints, 2D frequency fingerprints, and molecular properties. The detailed discussion for these descriptors can be found elsewhere (Todeschini and Consonni, 2000).

SiRMS Descriptors—2D Simplex Representation of Molecular Structure (SiRMS) descriptors (Muratov et al., 2010) (representing numbers of unique tetratomic fragments with fixed composition and topological structure) were generated by the HiT QSAR software (Kuz'min et al., 2008). At the 2D level, the connectivity of atoms in a simplex, atom type, and bond nature (single, double, triple, or aromatic) have been considered. SiRMS descriptors account not only for the atom type, but also for other atomic characteristics that may impact biological activity of molecules, *e.g.*, partial charge, lipophilicity, refraction, and an atom ability for being a donor/acceptor in hydrogen-bond formation (H-bond). For atom characteristics with continuous values (charge, lipophilicity, and refraction) the division of the entire value range into definite discrete groups has been carried out. The atoms have been divided into four groups corresponding to their *(i)* partial charge A −0.05<B 0<C 0.05<D, *(ii)* lipophilicity A −0.5<B 0<C 0.5<D, and *(iii)* refraction A 1.5<B 3<C 8<D. For H-bond characteristic, the atoms have been divided into three groups: A (acceptor of hydrogen in H-bond), D (donor of hydrogen in H-bond), and I (indifferent atom). The usage of sundry variants of differentiation of simplex vertexes (atoms) represents the principal feature of the SiRMS approach (Kuz'min et al., 2007). Detailed description of HiT QSAR and SiRMS can be found elsewhere (Kuz'min et al., 2008; Muratov et al., 2010).

QSAR modeling—The QSAR modeling workflow used in this study includes three major steps (Tropsha and Golbraikh, 2007; Tropsha, 2010): *(i)* data curation/preparation/analysis (selection of compounds and descriptors), *(ii)* model building, and *(iii)* model validation/selection. Here we followed a *5*-fold external cross-validation procedure: the full set of compounds with known experimental activity is randomly divided into five subsets of equal size; then one of these subsets (20% of all compounds) is set aside as an external validation set and the remaining four sets together form the modeling set (80% of the full set). This procedure is repeated five times allowing each of the five subsets to be used as external validation set. Models are built using the modeling set only, and it is important to emphasize that the compounds in corresponding external set (fold) are never employed either to build and/or select the models. Each modeling set is divided into many internal training and test sets; then models are built using compounds of each training set and applied to test set

compounds to assess their properties. The statistical metrics used to assess different aspects of model performance are available in the Supplementary Materials.

Best models were identified and selected according to acceptable threshold values (> 0.6) of Q^2_{ext} for the internal test sets (called out-of-bag set in Random Forest, *vide infra*). Then, selected models were applied to the external set compounds to predict their skin sensitization potential. This procedure was repeated five times to ensure that every compound was present once and only once in the corresponding external test set. Since the accuracy of each model is estimated only based on external set compounds, which are never used to derive, bias, or select models, this protocol ensures an objective estimation of the true external predictivity of the models. In addition, 1,000 rounds of Y-randomization were performed for each dataset to assure that the high accuracy of the models built with real data was not due to chance correlations.

Random Forest—Random Forest models were constructed according to the original RF algorithm (Breiman, 2001) using the CF software version 2.12 (Polishchuk et al., 2009). RF is an ensemble of single decision trees. Outputs of all trees are aggregated to obtain one final prediction. Each tree is grown as follows: *(i)* a bootstrap sample is produced from the whole set of N compounds to form a training set for the current tree. Compounds that are not in the training set of the current tree are placed in the out-of-bag (OOB) set (size of ~ N/3); *(ii)* the best split by CART algorithm (Breiman et al., 1984) among the *m* randomly selected descriptors from the entire pool in each node is chosen; *(iii)* each tree is then grown to the largest possible extent; there is no pruning. The predicted classification values are defined by majority voting for one of the classes. Thus, each tree predicts values for only those compounds that are not included in the training set of that tree (for OOB set only). Since RF possesses its own reliable statistical characteristics (based on OOB set prediction) which could be used for validation and model selection, no cross-validation was performed (Breiman, 2001). Thus, the final model is chosen by the lowest error for prediction of the OOB set. The local (tree) applicability domain approach (Artemenko et al., 2011) was used for all RF models developed in this study.

DERMWIN—The Dermal Permeability Coefficient Program (DERMWIN) is a module available in the EPI Suite package developed by the EPA's Office of Pollution Prevention Toxics and Syracuse Research Corporation and available at http://www.epa.gov/oppt/exposure/pubs/episuitedl.htm. The DERMWIN v.2.02 estimates chemicals' dermal permeability coefficient using the following equation 1 (USEPA, 2004):

$$logKp(cm/h) = -2.80 + 0.66 logK_{o/w} - 0.0056 MW, \quad (1)$$

where $K_{o/w}$ is the octanol/water partition coefficient and MW is the molecular weight of the chemical. Human skin permeability dataset (dataset D) was imported to DERMWIN as SMILES strings, and logKp was calculated and compared to the predictions of our models.

RESULTS

Relationship between human and rodent skin permeability coefficients

First, we searched for the subset of chemicals that had both human and rodent skin permeability data available to verify whether we could increase the dataset size by merging these data. We found 34 compounds that have both human and rodent experimental data. As is obvious from Figure 1, the correlation between human and rodent data was not high enough to merge these datasets.

A linear regression of the logKp values yielded R^2=0.44 only. Therefore, we decided that the data for humans and rodents were definitely not compatible and built species-specific QSAR models. Human-based and rodent-based QSAR models for skin permeability were developed using the same protocol used for skin sensitization as described in the companion paper (Alves et al., 2014). The results (Table 1) suggest that continuous models for both datasets afforded high predictive accuracy as estimated by Q^2_{ext} values. However, the quality and coverage of human-based models 12 and 14 were higher than the same characteristics of the best rodent-based model 21. Thus, models 12 and 14 were retained for *in silico* screening of chemical libraries of concern to identify potential human skin sensitizers. Additional statistical characteristics of the models are present in Supplementary Materials (Table S2).

Relationship between skin sensitization and permeability

We have explored whether there are intrinsic relationships between skin sensitization and skin permeability of chemicals. Since rodent data did not correlate well with human data, only human skin permeability data were employed for this analysis. We identified a subset of 20 compounds for which both skin sensitization and skin permeability data were known. These experimental data were compiled from several studies (Chauhan and Shakya, 2010; ICCVAM, 2009; Jaworska et al., 2011). As shown in Table 2, there is no direct relationship between the two endpoints. The ranges of logKp values for sensitizers (logKp = −3.62 – −1.28) *vs*. non-sensitizers (logKp = −3.05 – −1.6) confirm this finding. For instance, five non-sensitizers have low permeability coefficient while logKp values for two other non-sensitizers are high. There is a common understanding that a chemical should penetrate the skin to cause sensitization (penetration is generally regarded as the first step in the skin sensitization AOP (Karlberg et al., 2008; OECD, 2012)). However, we found examples when a relatively weak penetrant can still be a strong sensitizer (*e.g*., p-Phenylenediamine, logKp = −3.62), whereas a strong penetrant, can be a non-sensitizer (*e.g*., octanoic acid, logKp = −1.60). This observation suggests that skin permeability and skin sensitization are generally decoupled processes and therefore it may be possible to modify chemicals such as to affect permeability without affecting sensitization, and vice versa.

Cluster analysis of human skin permeability dataset

We have performed a cluster analysis of the human skin permeability set (dataset D) following the same method used for the skin sensitization dataset A (part I in (Alves et al., 2014)). Due to the poor concordance between human and rodent data, and a heightened interest in human skin permeability data, this analysis was only conducted using dataset D.

As illustrated in Figure 2, the dendrogram for the dataset D is not as extensive as for skin sensitization dataset A (Alves et al., 2014), *i.e.*, compounds in the skin penetration dataset D are more structurally similar than those in the skin sensitization dataset A. Short-chain alcohols have similar logKp values (ca. −3.0) except for ethylene glycol (logKp = −4.07). For glycol ethers with short carbon chains (cluster b in Figure 2), the decrease of permeability is likely related to the chain length (2-(2-methoxyethoxy)ethanol (logKp= −3.69) < 2-(2-ethoxyethoxy)ethanol (logKp= −3.88) < 2-(2-butoxyethoxy)ethanol) (logKp= −4.45). To confirm this tendency, we have calculated permeability for two additional glycol ethers. We found that 2-(2-propoxyethoxy)ethanol (logKp$_{pred}$ = −4.05), which had an logKppred intermediary between 2-(2-ethoxyethoxy)ethanol and 2-(2-butoxyethoxy)ethanol, followed the trend but our predictions indicated a slight increase in permeability for 2-(2-(pentyloxy)ethoxy)ethanol (logKp$_{pred}$ = −3.95). Then, with further increase of the chain length, this effect reached saturation resulting in a highly similar permeability for the next three compounds: 2-(2-(hexyloxy)ethoxy)ethanol (logKp$_{pred}$ = −3.88), 2-(2-(heptyloxy)ethoxy)ethanol (logKp$_{pred}$ = −3.87), and 2-(2-(octyloxy)ethoxy)ethanol (logKp$_{pred}$ −3.89). Two compounds, 2-(2-(nonyloxy)ethoxy)ethanol and 2-(2-(decyloxy)ethoxy)ethanol were outside of the AD for the consensus model 5.

As a result of cluster analysis, we found three compounds that had different activity annotations from the rest of the cluster members and were therefore suspicious. These compounds included a barbiturate (amylobarbital) and two steroid derivatives of hydrocortisone (hydrocortisone methylsuccinate, and hydrocortisone succinamate). Our procedure for evaluating suspicious compounds was based on the permeability prediction for these compounds using developed models and a search for additional experimental data to verify their human skin permeability coefficients. Unfortunately, we were unable to find any confirmatory data in the literature to prove or refute the values of permeability coefficients for these substances. According to our consensus model 5, amylobarbital had a predicted logKp = −3.56 (experimental logKp= −2.64) similar to other barbitals: barbital (logKp = −3.95), phenobarbital (logKp = −3.34), and butobarbital (logKp = −3.71), which made us less confident in the accuracy of the reported permeability of this compound. Within steroids (cluster c in Figure 2), progesterone was the strongest penetrant (logKp = −1.89), whereas the addition of one hydroxyl group in position 17 of the steroid scaffold significantly decreased the permeability (hydroxyprogesterone logKp = −3.22). The hydroxyl group in position 17 appears to be bad for steroid permeability, since testosterone (logKp = −3.40) contains the hydroxyl group but lacks the ethoxy group. We removed both hydroxyl and ethoxy group from position 17, and the resulting compound, 3-oxo-delta4-steroid, had a predicted logKp$_{pred}$ = −2.28; adding a methyl group to the same position resulted in a logKp$_{pred}$ = −2.28, which was higher than the permeability of hydroxyprogesterone and testosterone, but lower than that of progesterone.

The presence of other substituents and positions of double bonds seemed to be less important. The effects of the substituent in position 21 on compound permeability seem to depend on its size: it decreases when a methyl substituent (methylsuccinate; logKp = −3.68) is replaced by a larger group (succinamate substituent; logKp = −4.59) but dramatically increases as shown in the four hydrocortisone derivatives (hydrocortisone hexanoate,

hydrocortisone octanoate, hydrocortisone methylpimelate, and hydrocortisone hemipimelate) that have a long chain in this position (see Table S4). In the beginning of the homological series of compounds, the permeability of carboxylic acids (cluster a in Figure 2) increased proportionally to the length of the hydrocarbon chain, *i.e.*, acetic acid (logKp = −3.21) < propionic acid (logKp = −2.74) < butanoic acid (logKp = −2.90) < pentanoic acid (logKp = −2.70) < hexanoic acid (logKp = −1.85) < heptanoic acid (logKp = −1.70) < octanoic acid (logKp = −1.60). Then, similarly to glycol ethers from cluster b, the effect reached saturation and predicted logKp values for five next carboxylic acids became very similar: nonanoic acid (logKp$_{pred}$ = −1.68), decanoic acid (logKp$_{pred}$ = −1.68), undecanoic acid (logKp$_{pred}$ = −1.71), dodecanoic acid (logKp$_{pred}$ = −1.82), and tridecanoic acid (logKp$_{pred}$ = −1.84).

QSAR modeling of skin permeability

We have employed a dataset (Chauhan and Shakya, 2010) comprising 211 compounds (dataset D). Several structural duplicates were identified during the curation process. The complete list of 17 duplicate pairs and two sets of triplicates is shown in Table S3. Skin permeability values for all pairs of duplicates were the same or almost the same. This observation also means that the statistical performance of the models developed in an earlier study (Chauhan and Shakya, 2010) may have been over-optimistic because, as shown previously (Fourches et al., 2010), the presence of duplicates with identical activity annotations in both training and test sets generally leads to an overestimation of the model's predictivity. Here, we built species-specific QSAR models of skin permeability using the same protocol used for the skin sensitization study (Alves et al., 2014). The results shown in Table 1 suggest that continuous models with high accuracy for both datasets have been generated; however, the quality and coverage of human-based models 12 and 14 were higher than the same characteristics of rodent-based model 21. The former models have been retained for *in silico* screening of chemical libraries.

Comparison of developed QSAR models *vs.* DERMWIN permeability predictions

The DERMWIN module available as part of the EPI Suite software has been employed in several recent studies (Fong and Tong, 2012; Fong et al., 2014; Zhang et al., 2013) for evaluating skin permeability. For this reason, we decided to compare our models with that implemented in this software. DERMWIN uses a linear equation that relates the permeability coefficient with the chemical's octanol/water partition coefficient and molecular weight. However, statistical characteristics of this model suggested that this correlation was not adequate to predict human permeability. Although our consensus model 12 had a lower coverage of chemical space than DERMWIN (77% vs. 100%), it significantly outperformed DERMWIN in predictivity (Q^2_{ext} = 72% *vs.* 43%, respectively) (see Table 1). When the AD restriction was removed and the predictive accuracy of the models was evaluated using the same set of compounds (143 compounds), the comparison was in favor of our model (Q^2_{ext} = 71% *vs.* 66%, respectively) (see Table 3).

DISCUSSION

Can rodent skin permeability data be used to predict human permeability?

As presented in the Results section, we found 34 compounds that had experimental permeability coefficient data for both human and rodents. Although the overall correlation between skin permeability measured in these two species was not high, we wished to determine whether the model obtained on rodent dataset E could be used to predict skin permeability in humans. Virtual screening of the human dataset D using the model developed with the rodent data resulted in a reasonably high predictivity (Q^2_{ext} = 0.77, RMSE = 1.27) but only for a very small fraction (19%) of compounds. Based on these results, one can assume that the rodent model could be used for predicting human skin permeability for at least some compounds within the conservative AD, but this observation should be considered with caution since it does not hold true for a larger data set. Additional experimental results for greater number of compounds are needed for more definitive conclusions.

Relationships between skin sensitization potential and skin permeability based on computed data

Given the small number of compounds with known experimental values of both skin sensitization and permeability, we decided to apply our selected QSAR models to cross-predict these properties. The use of QSAR-imputed permeability and sensitization data allowed us to examine the possible relationships between the two endpoints for a much larger set of compounds. Skin sensitization model **5** (see the companion paper (Alves et al., 2014)) was used for predicting skin sensitization for 186 compounds in the human skin permeability dataset (D), whereas skin permeability model **5** (Table 1) was used for calculating permeability coefficients for 387 compounds in the skin sensitization dataset retrieved from ICCVAM (254 compounds from dataset A + 133 sensitizers from dataset B) (see (Alves et al., 2014)). Then, we combined all predictions for compounds with unavailable experimental data (see Table S5). Compounds outside of AD either for model 5 or 12 were removed from the analysis. In the end, 283 compounds that had information (experimental or predicted) for both endpoints were considered for the analysis. As one can see in Figure S1, all but one of the weakest penetrants (logKp < −5) were non-sensitizers. The only exception was 1-dodecyl glycidyl ether, which could have been mispredicted by model **5**. It was hypothesized recently that compound accumulation in the skin is contributing more significantly to skin sensitization than its actual permeability through the skin layer (Jaworska et al., 2013). Our results (see Figure S1) show that easily-penetrating compounds may be both sensitizers and non-sensitizers. For compounds with low permeability we made the same observation. These findings re-confirm the absence of the global correlation between skin permeability and sensitization.

Aforementioned results contradict a common view of skin permeation as the first step in the adverse outcome pathway leading to skin sensitization (OECD, 2012). To investigate whether the concordance between skin sensitization and permeability could be confined to certain chemical classes, we have performed a cluster analysis for the entire dataset of 283 compounds with imputed data. Careful analysis of the relationships between the two

endpoints within clusters of structurally similar compounds showed that the observation linking the high permeability and skin sensitization potential could be made only for one cluster out of 34. This specific Cluster 1 is formed by 11 compounds represented by four barbitals (barbital, amylobarbital, butobarbital, and phenobarbital) and seven other drugs (scopolamine, ibuprofen, atropine, griseofulvin, cyclamen aldehyde, p-tert-butyl-alpha-ethyl-hydrocinnamal, and sodium-3,3,5-trimethyl-benzenesulfonate). The comparison between permeability and sensitization for compounds in this cluster is shown in Figure S2 and Table S6. The analysis revealed that seven compounds are associated with $\log Kp < -2.5$ and are non-sensitizers, whereas the four other compounds with $\log Kp > -2.5$ are sensitizers.

No distinct trend could be observed for the remaining clusters. Although, as noticed above, skin permeability is widely considered to be mechanistically implicated in skin sensitization, we have found no evidence that high permeability implies high skin sensitization potency and vice versa. Some other authors reached the same conclusions. For instance, the analysis of mechanism-based QSARs demonstrates no role for skin permeability in determining potency (Roberts and Aptula, 2008). Similarly, in another recent study (Jaworska et al., 2013) the authors hypothesizes that the accumulation of a chemical in the epidermis layer of the skin is much more related to the skin sensitization than the actual penetration properties. In another study (Roberts and Patlewicz, 2010), it was hypothesized that skin sensitizers reach the viable epidermis by-passing skin permeability route via shunt pathways, since bad penetrants have shown to be sensitizers when applied by intradermal injection.

Another important benefit of this cheminformatics analysis is the use of structural rules established from the interpretation of developed QSAR models, SAR and cluster analysis, in order to design new compounds with improved permeability and sensitization characteristics. We illustrated this approach using an example of putative stepwise structural optimization for permeability of pentanoic acid, considering experimental data and predictions using developed models (Figure 3). Starting from this compound with a relatively low permeability ($\log Kp = -2.7$), several transformation steps can increase its permeability more than 10-fold and convert it to n-heptanol ($\log Kp = -1.50$). n-heptanol is predicted as a sensitizer (which is confirmed by its Material Safety Data Sheet (OXEA, 2013) but it can be transformed to octanoic acid that has a similar permeability ($\log Kp = -1.60$) as n-heptanol but lacks its sensitization potential.

CONCLUSIONS

We have compiled, curated, and integrated the largest publicly available datasets of skin permeability for diverse chemicals. The analysis of the experimental data for compounds containing both skin sensitization and permeability data indicated that, with a few exceptions, there is no overall concordance between these two endpoints, *i.e.*, weak penetrants could be strong sensitizers and *vice versa*. Although sensitizers have to penetrate the skin layer, the permeability coefficient is not a determinant of the skin sensitization potential. Cluster analysis also helped us to highlight the high consistency of experimental Kp data reported in the literature.

We have built statistically significant and externally predictive QSAR models of skin permeability that can be used to predict the permeability of untested compounds through the skin. Comparison of the developed consensus model with DERMWIN software showed that our model significantly outperformed DERMWIN in predictivity (Q^2_{ext} = 72% and 43%, respectively) but at the expense of some lost in coverage (77% vs. 100%); when applied to the same set of compounds as used for DERMWIN (ignoring the AD for our models) the performance of our model is still higher (Q^2_{ext} = 71% vs. 66%). The compiled datasets and all the models developed in this study have been made publicly available at the Chembench Web Portal (http://chembench.mml.unc.edu).

The use of skin permeability and sensitization values imputed by our QSAR models allowed us to examine the relationships between these two endpoints utilizing a significantly expanded set consisting of 283 compounds. The results indicated that there is still no overall concordance between these endpoints. This phenomenon could be explained by attributing higher impact to compound accumulation in the skin on sensitization potential, rather than permeability through the skin layer. Further investigation into this observation would be facilitated via building predictive models of compound accumulation in the skin, and incorporating considerations of protein binding and reactivity.

In conclusion, the skin permeability models developed in this study could be useful for risk assessment purposes in addition to the assessment of skin sensitization; for instance, the knowledge how much of an applied dose will penetrate the skin may be helpful to estimate systemic exposure to topically applied chemicals. Moreover, the lack of correlation between skin permeability and skin sensitization established in this study suggests a possibility of rational design of compounds with the desired high permeability but low skin sensitization potential that may be of value to the cosmetics industry.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

ABBREVIATIONS

AD	Applicability domain
AOP	Adverse outcome pathway

ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
Kp	Permeability coefficient
LLNA	Local lymph node assay
LU	Logarithmic unit
MAE	Mean absolute error
OECD	Organization for Economic Co-operation and Development
(Q)SAR	(Quantitative) structure-activity relationship
RMSE	Root mean squared error

REFERENCES

Abraham MH, Chadha HS, Martins F, Mitchell RC, Bradbury MW, Gratton JA. Hydrogen bonding part 46: a review of the correlation and prediction of transport properties by an lfer method: physicochemical properties, brain penetration and skin permeability. Pestic. Sci. 1999; 55:78–88.

Alves VM, Muratov EN, Fourches D, Strickland J, Kleinstreuer N, Andrade CH, Tropsha A. Predicting chemically-induced skin reactions. Part I: QSAR models of skin sensitization and their application to identify potentially hazardous compounds. Toxicol. Appl. Pharmacol. 2014 Submitted.

Artemenko AG, Muratov EN, Kuz'min VE, Muratov NN, Varlamova EV, Kuz'mina AV, Gorb LG, Golius A, Hill FC, Leszczynski J, Tropsha A. QSAR analysis of the toxicity of nitroaromatics in Tetrahymena pyriformis: structural factors and possible modes of action. SAR QSAR Environ. Res. 2011; 22:575–601. [PubMed: 21714735]

Barratt MD. Quantitative structure-activity relationships for skin permeability. Toxicol. In Vitro. 1995; 9:27–37. [PubMed: 20650060]

Berge W. A simple dermal absorption model: derivation and application. Chemosphere. 2009; 75:1440–1445. [PubMed: 19304310]

Breiman, L.; Friedman, JH.; Olshen, RA.; Stone, CJ. Classification and Regression Trees. Belmont: Wadsworth Publishing; 1984.

Breiman LEO. Random Forests. Mach. Learn. 2001; 45:5–32.

Brown MB, Lau C-H, Lim ST, Sun Y, Davey N, Moss GP, Yoo S-H, De Muynck C. An evaluation of the potential of linear and nonlinear skin permeation models for the prediction of experimentally measured percutaneous drug absorption. J. Pharm. Pharmacol. 2012; 64:566–577. [PubMed: 22420662]

Chauhan P, Shakya M. Role of physicochemical properties in the estimation of skin permeability: in vitro data assessment by Partial Least-Squares Regression. SAR QSAR Environ. Res. 2010; 21:481–494. [PubMed: 20818583]

Chen L, Han L, Lian G. Recent advances in predicting skin permeability of hydrophilic solutes. Adv. Drug Deliv. Rev. 2013; 65:295–305. [PubMed: 22580335]

Chen L, Lian G, Han L. Modeling transdermal permeation. Part I. Predicting skin permeability of both hydrophobic and hydrophilic solutes. AIChE J. 2010; 56:1136–1146.

Chen LL, Lian GG, Han LL. Prediction of human skin permeability using artificial neural network (ANN) modeling. Acta Pharmacol. Sin. 2007; 28:591–600. [PubMed: 17376301]

Cronin MT, Dearden JC, Moss GP, Murray-Dickson G. Investigation of the mechanism of flux across human skin in vitro by quantitative structure-permeability relationships. Eur. J. Pharm. Sci. 1999; 7:325–330. [PubMed: 9971916]

Downs, GM.; Barnard, JM. Clustering Methods and Their Uses in Computational Chemistry. In: Lipkowitz, KB.; Boyd, DB., editors. Reviews in Computational Chemistry. Hoboken: John Wiley & Sons, Inc.; 2003. p. 1-40.

Flynn, GL. Physicochemical determinants of skin absorption. In: Gerrity, TR.; Henry, CJ., editors. Principles of Route-to-Route Extrapolation for Risk Assessment. New York, NY: Elsevier; 1990. p. 93-117.

Fong P, Tong HHY. In silico prediction of the cosmetic whitening effects of naturally occurring lead compounds. Nat. Prod. Commun. 2012; 7:1287–1294. [PubMed: 23156992]

Fong P, Tong HHY, Chao CM. In silico prediction of tyrosinase and adenylyl cyclase inhibitors from natural compounds. Nat. Prod. Commun. 2014; 9:189–194. [PubMed: 24689287]

Fourches D, Muratov E, Tropsha A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. J. Chem. Inf. Model. 2010; 50:1189–1204. [PubMed: 20572635]

Hostýnek JJ, Magee PS. Fragrance allergens: Classification and ranking by QSAR. Toxicol. In Vitro. 1997; 11:377–384. [PubMed: 20654324]

ICCVAM. The reduced murine local lymph node assay: an alternative test method using fewer animals to assess the allergic contact dermatitis potential of chemicals and products. 2009. [WWW Document]. URL http://ntp.niehs.nih.gov/iccvam/docs/immunotox_docs/LLNA-LD/TMER.pdf (accessed 2.20.12).

Jaworska J, Dancik Y, Kern P, Gerberick F, Natsch A. Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. J. Appl. Toxicol. 2013; 33:1353–1364. [PubMed: 23670904]

Jaworska J, Harol A, Kern PS, Gerberick GF. Integrating non-animal test information into an adaptive testing strategy - skin sensitization proof of concept case. ALTEX. 2011; 28:211–225. [PubMed: 21993957]

Johansen, JD.; Frosch, PJ.; Menné, T. Allergic contact dermatitis in humans: experimental and quantitative aspects. In: Johansen, JD.; Frosch, PJ.; Lepoittevin, J-P., editors. Contact Dermatitis. Berlin: Springer; 2011. p. 241-251.

Karlberg A-T, Bergström MA, Börje A, Luthman K, Nilsson JLG. Allergic contact dermatitis - formation, structural requirements, and reactivity of skin sensitizers. Chem. Res. Toxicol. 2008; 21:53–69. [PubMed: 18052130]

Kuz'min VE, Artemenko aG, Muratov EN. Hierarchical QSAR technology based on the Simplex representation of molecular structure. J. Comput. Aided. Mol. Des. 2008; 22:403–421. [PubMed: 18253701]

Kuz'min VE, Artemenko AG, Muratov EN, Volineckaya IL, Makarov VA, Riabova OB, Wutzler P, Schmidtke M. Quantitative structure-activity relationship studies of [(biphenyloxy)propyl]isoxazole derivatives. Inhibitors of human rhinovirus 2 replication. J. Med. Chem. 2007; 50:4205–4213. [PubMed: 17665898]

Lien EJ, Gao H. QSAR analysis of skin permeability of various drugs in man as compared to in vivo and in vitro studies in rodents. Pharm. Res. 1995; 12:583–587. [PubMed: 7596996]

MacKay C, Davies M, Summerfield V, Maxwell G. From pathways to people: applying the adverse outcome pathway (AOP) for skin sensitization to risk assessment. ALTEX. 2013; 30:473–486. [PubMed: 24173169]

Magnusson BM, Anissimov YG, Cross SE, Roberts MS. Molecular size as the main determinant of solute maximum flux across the skin. J. Invest. Dermatol. 2004; 122:993–999. [PubMed: 15102090]

Mercier D. Clustering large datasets. Electron. Rev. - Linacre Coll. 2003

Moss GP, Cronin MTD. Quantitative structure-permeability relationships for percutaneous absorption: re-analysis of steroid data. Int. J. Pharm. 2002; 238:105–109. [PubMed: 11996814]

Moss GP, Sun Y, Wilkinson SC, Davey N, Adams R, Martin GP, Prapopopolou M, Brown MB. The application and limitations of mathematical modelling in the prediction of permeability across mammalian skin and polydimethylsiloxane membranes. J. Pharm. Pharmacol. 2011; 63:1411–1427. [PubMed: 21988422]

Muratov EN, Artemenko AG, Varlamova EV, Polischuk PG, Lozitsky VP, Fedchuk AS, Lozitska RL, Gridina TL, Koroleva LS, Sil'nikov VN, Galabov AS, Makarov VA, Riabova OB, Wutzler P, Schmidtke M, Kuz'min VE. Per aspera ad astra: application of Simplex QSAR approach in antiviral research. Future Med. Chem. 2010; 2:1205–1226. [PubMed: 21426164]

OECD. OECD principles for the validation, for regulatory purposes, of (Quantitative) Structure-Activity Relationship models. 2004 [WWW Document]. URL http://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf (accessed 4.11.13).

OECD. The adverse outcome pathway for skin sensitisation initiated by covalent binding to proteins part 1: scientific evidence. OECD Enviroment, Heal. Safaty Publ. 2012 [WWW Document]. URL http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en (accessed 4.11.12).

OXEA. n-Heptanol. 2013 [WWW Document]. URL http://www.oxea-chemicals.com/download/wercs/MTA5MDAjZW4jcHMjYXVzIzEzNjg2MTA1NzIwMDAjb3hlYSMxI0Q=/10900-en-ps-us.pdf (accessed 11.18.14).

Patel H, ten Berge W, Cronin MTD. Quantitative structure-activity relationships (QSARs) for the prediction of skin permeation of exogenous chemicals. Chemosphere. 2002; 48:603–613. [PubMed: 12143935]

Polishchuk PG, Muratov EN, Artemenko AG, Kolumbin OG, Muratov NN, Kuz'min VE. Application of random forest approach to QSAR prediction of aquatic toxicity. J. Chem. Inf. Model. 2009; 49:2481–2488. [PubMed: 19860412]

Potts RO, Guy RH. Predicting skin permeability. Pharm. Res. 1992; 9:663–669. [PubMed: 1608900]

Potts RO, Guy RH. A predictive algorithm for skin permeability: the effects of molecular size and hydrogen bond activity. Pharm. Res. 1995; 12:1628–1633. [PubMed: 8592661]

Roberts DW, Aptula AO. Determinants of skin sensitisation potential. J. Appl. Toxicol. 2008; 28:377–387. [PubMed: 17703504]

Roberts DW, Patlewicz GY. Updating the skin sensitization in vitro data assessment paradigm in 2009--a chemistry and QSAR perspective. J. Appl. Toxicol. 2010; 30:286–288. discussion 289. [PubMed: 20131408]

Todeschini, R.; Consonni, V. Handbook of Molecular Descriptors, Methods and Principles in Medicinal Chemistry. Weinheim, Germany: Wiley-VCH; 2000.

Tropsha A. Best practices for QSAR model development, validation, and exploitation. Mol. Inform. 2010; 29:476–488.

Tropsha A, Golbraikh A. Predictive QSAR modeling workflow, model applicability domains, and virtual screening. Curr. Pharm. Des. 2007; 13:3494–3504. [PubMed: 18220786]

USEPA. Risk Assessment Guidance for Superfund (RAGS), Volume I: Human Health Evaluation Manual (Part E Supplemental Guidance for Dermal Risk Assessment) Interim. 2004. [WWW Document]. URL http://www.epa.gov/oswer/riskassessment/ragse/index.htm (accessed 5.13.13).

Varnek A, Fourches D, Horvath D, Klimchuk O, Gaudin C, Vayer P, Solov'ev V, Hoonakker F, Tetko I, Marcou G. ISIDA - Platform for Virtual Screening Based on Fragment and Pharmacophoric Descriptors. Curr. Comput. Aided-Drug Des. 2008; 4:191–198.

Varnek A, Fourches D, Sieffert N, Solov'ev VP, Hill C, Lecomte M. QSPR Modeling of the Am III /Eu III Separation Factor: How Far Can we Predict? Solvent Extr. ion Exch. 2007; 25:1–26.

Xia, XR. Quantitative Structure-Permeability Relationships. In: Riviere, JE., editor. Comparative Pharmacokinetics: Principles, Techniques, and Applications. John Wiley & Sons, Inc.; 2011. p. 27-38.

Zhang F, Zhang J, Tong C, Chen Y, Zhuang S, Liu W. Molecular interactions of benzophenone UV filters with human serum albumin revealed by spectroscopic techniques and molecular modeling. J. Hazard. Mater. 2013; 263(Pt 2):618–626. [PubMed: 24231334]

Highlights

- We compiled the largest publicly-available skin permeability dataset.

- Predictive QSAR models were developed for skin permeability.

- No concordance between skin sensitization and skin permeability has been found.

- Structural rules for optimizing sensitization and penetration were established.
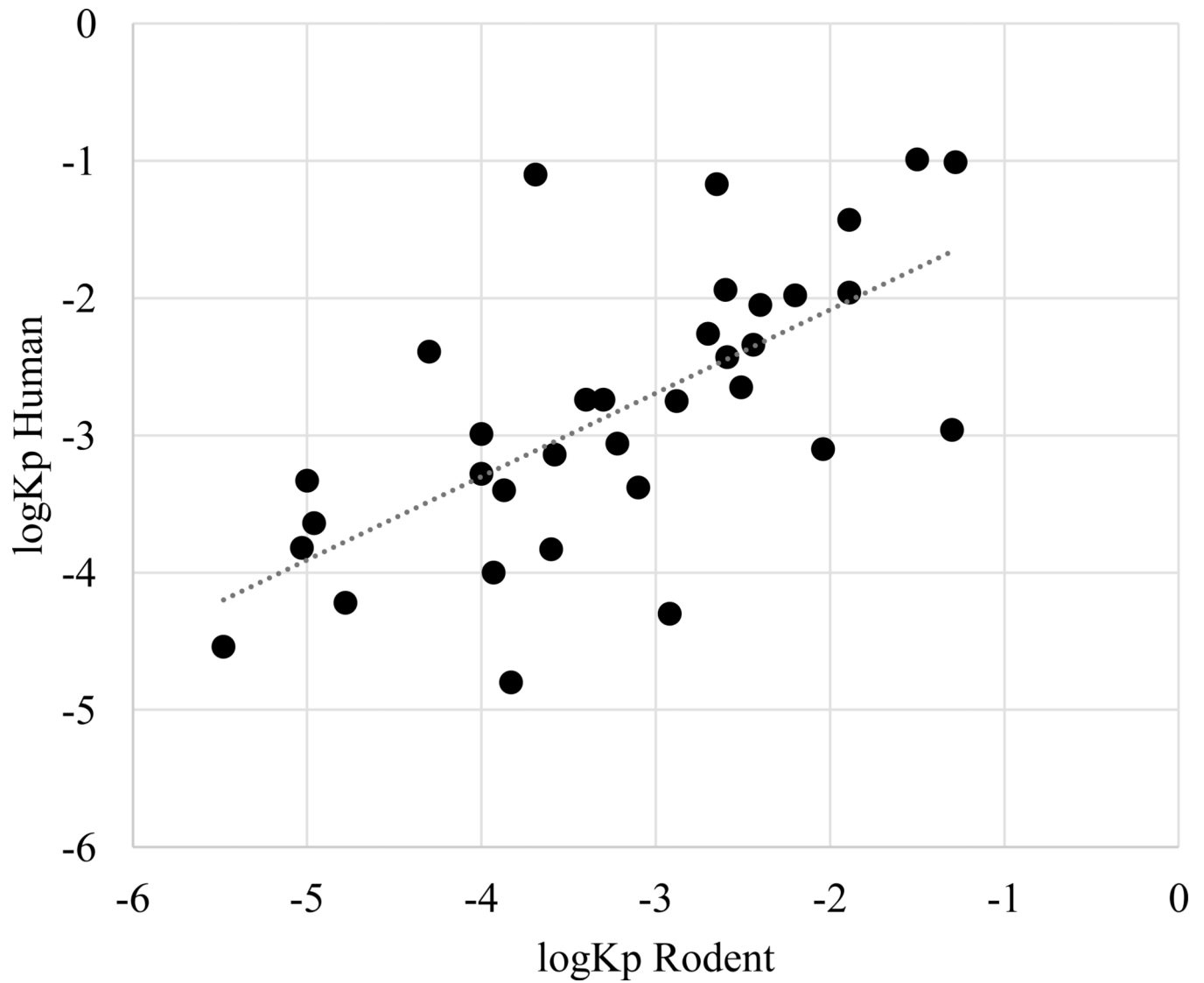
Figure 1.
Human *vs*. rodent skin permeability correlation ($R^2 = 0.44$, $logKp_{Human} = 0.72*logKp_{Rodent} - 1.17$).
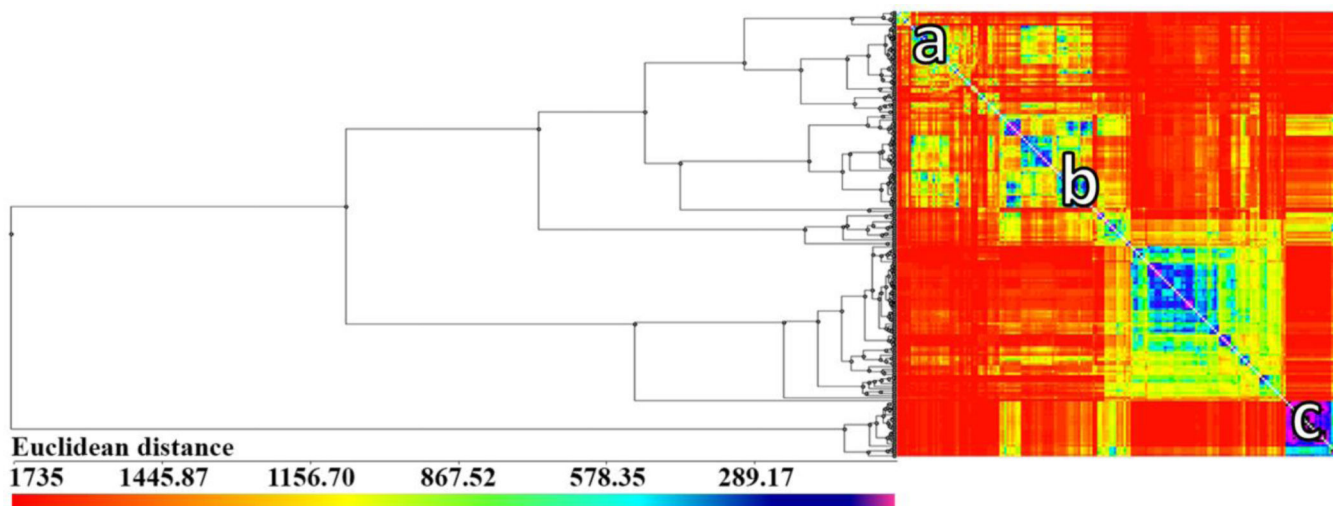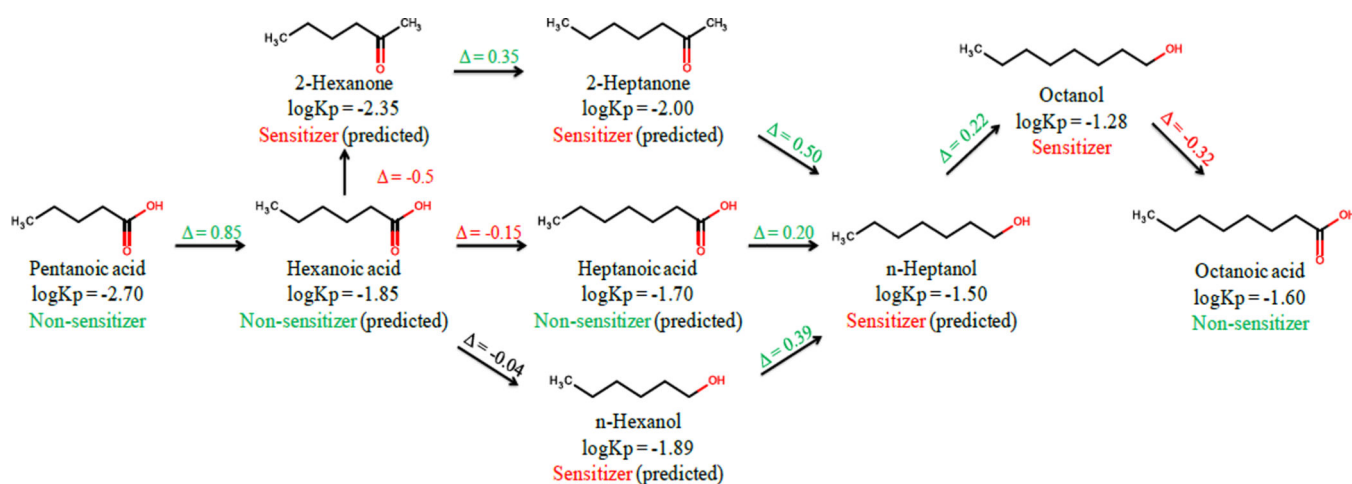
Figure 2.
Cluster analysis of the human skin permeability dataset D: dendrogram and heat map of the distance matrix ordered based on structural similarity (blue/violet = similar; yellow/red = dissimilar). The following clusters are noted: (a) carboxylic acids, (b) glycol ethers, and (c) steroids.

Figure 3.
Example of a structural transformation of sensitizer n-octanol with low permeability to non-sensitizer octanoic acid with improved permeability. Desired change of property is highlighted by green, undesired – by red; Δ = logKp$_{parent}$ − logKp$_{child}$.

Table 1

Statistical characteristics of QSAR models for skin permeability assessed by 5-fold external cross-validation.

	Models	Q^2_{ext}	RMSE	MAE	Coverage
	Model 1 - SiRMS	0.69	0.52	0.40	0.72
	Model 2 - SiRMS no AD*	0.50	0.73	0.53	1.00
	Model 3 - Dragon	0.73	0.46	0.35	0.68
	Model 4 - Dragon no AD*	0.55	0.70	0.49	1.00
HUMAN	Model 5 - Consensus	0.72	0.49	0.38	0.77
	Model 6 - Consensus no AD*	0.55	0.69	0.50	1.00
	Model 7 - Consensus Rigor	0.73	0.45	0.34	0.63
	DERMWIN	0.43	0.78	0.49	1.00
	Model 8 - SiRMS	0.57	0.67	0.50	0.71
	Model 9 - SiRMS no AD*	0.35	0.82	0.61	1.00
	Model 10 - Dragon	0.33	0.75	0.56	0.69
	Model 11 - Dragon no AD*	0.32	0.84	0.63	1.00
RODENT	Model 12 - Consensus	0.41	0.77	0.56	0.86
	Model 13 - Consensus no AD*	0.38	0.80	0.61	1.00
	Model 14 - Consensus Rigor	0.61	0.58	0.43	0.53

Notes: Models 1 to 7: Human-based skin permeability models. Models 8 to 13: Rodent-based skin permeability models. RSME: root mean square error; MAE: mean absolute error

*
Applicability Domain was not considered in these models.

Table 2

Comparison of experimental data on skin sensitization and human skin permeability

Chemical compound	LLNA Result	logKp
1,6-Hexanediol diglycidyl ether [1–3]	Sensitizer (moderate)	−3.87
Resorcinol [1–3]	Sensitizer (moderate)	−3.62
p-Phenylenediamine [1–3]	Sensitizer (strong)	−3.62
2-Nitro-4-phenylenediamine [1–3]	Sensitizer (strong)	−3.30
Isopropyl alcohol [1–3]	Non-sensitizer	−3.05
Pyridine [1–3]	Sensitizer (weak)	−2.74
Clotrimazole [1–3]	Sensitizer (moderate)	−2.70
Methyl acrylate [1–3]	Sensitizer (weak)	−2.68
Formaldehyde [1–3]	Sensitizer (strong)	−2.65
Aniline [1–3]	Sensitizer (weak)	−2.65
n-Butanol [1–3]	Non-sensitizer	−2.60
Methyl acrylic acid [1–3]	Sensitizer (weak)	−2.58
Ethyl acrylate [1–3]	Sensitizer (weak)	−2.39
Salicylic acid [1–3]	Non-sensitizer	−2.20
Coumarin [1–3]	Non-sensitizer	−2.04
Methyl 4-hydroxybenzoate [1–3]	Non-sensitizer	−2.04
Butyl acrylate [1–3]	Sensitizer (weak)	−2.00
Octanoic acid [1–3]	Non-sensitizer	−1.60
1-Naphthol [1–3]	Sensitizer (moderate)	−1.55
n-Octanol [1, 2]	Sensitizer (moderate)	−1.28

Notes: [1] (Chauhan and Shakya, 2010); [2] (ICCVAM, 2009); [3] (Jaworska et al., 2011); Kp: permeability coefficient.

Table 3

Comparison of the performance of consensus QSAR model with that of DERMWIN on the set of 143 compounds.

Models	Q^2_{ext}	RMSE	MAE
Model 5 -Consensus	0.72	0.49	0.38
DERMWIN	0.66	0.53	0.37