Predicting skin permeability of chemical substances using a quantitative structure-activity relationship

Yen-Ching CHANG\textsuperscript{a}, Chen-Peng CHEN\textsuperscript{a,*}, Chan-Cheng CHEN\textsuperscript{b}

\textsuperscript{a}Department of Occupational Safety and Health, China Medical University, No. 91 Hsueh-Shih Road, Taichung 40402, Taiwan, China
\textsuperscript{b}Department of Safety, Health and Environmental Engineering, National Kaohsiung First University of Science and Technology, No. 2 Jhuoyue Road, Nanzih, Kaohsiung City 81164, Taiwan, China

Abstract

The health hazard arising from skin exposure to toxic chemical was conventionally evaluated by determining the skin permeation coefficient ($K_p$) of chemical in the stratum corneum of human or animal cadaver skin. However, limited by constraints inherent in the technique, to date the $K_p$ evaluation has been performed only insufficiently. This study defined a quantitative structure-activity relationship (QSAR) for $K_p$ prediction and for characterization of, at a molecular level, the physicochemical properties involved in transdermal transport of chemical. One hundred and fifty-eight chemical substances of known $K_p$ determined in vitro using human skin were selected in the QSAR development. The final QSAR consisted of four molecular descriptors, including those describing electrostatic interactions between electric quadrupoles of van der Waals forces, octanol-water partitioning of solute, similarity in antineoplastic property, and frequency of carbon-nitrogen bonding at a constant topological distance. The four-descriptor multiple linear regression model fit the observed $K_p$ data with a $R^2$ of 0.828 and a mean percentage error of 18.8%, suggesting a capacity of this QSAR to serve as an alternative source of $K_p$ information and a potential tool of dermal hazard characterization, particularly when experimentally determined $K_p$ are not readily available.

Keywords: Quantitative structure-activity relationship; skin permeability; industrial chemical

1. Introduction

The exposure of the skin to chemical of toxic potency in the workplace has become an important issue in industrial and regulatory toxicology. With the inhalational exposure to airborne toxicants being effectively controlled, the systemic uptake of chemical by route of skin absorption poses an increasing threat to the workers' health. The US National Institute for Occupational Safety and Health estimated that 42% of the US workers risked the exposure of their skin to toxic industrial chemical; the US Bureau of Labor Statistics based on the findings of their annual survey also pointed out that over 13 millions of the US workforce were everyday in contact of chemical capable of provoking systemic toxicity by route of skin absorption [1]. The health hazard arising from dermal absorption of chemical was, conventionally, evaluated by determining the skin permeation coefficient ($K_p$) of chemical in the stratum corneum, i.e. the rate of permeation through the outermost layer of the epidermal skin. The $K_p$ was typically determined in vitro using human or animal cadaver skin. However, limited by the constraints in the experimental techniques, this evaluation had only been insufficiently performed.

As an alternative, since the 90s the quantitative structure-activity relationship (QSAR) has received attention as a strategy in hazard evaluation, and various QSAR models have been attempted to provide a viable means of $K_p$ prediction for...
industrial chemicals of varying physicochemical properties. As the technique advances, the application of QSAR in \(Kp\) modeling continues to evolve to provide algorithms of greater predictive capacity and capability. Lian et al. [2] reviewed a list of mechanistic and empirical QSARs widely accepted for and applied in \(Kp\) prediction, and indicated that the mechanistic and empirical \(Kp\)-predicting QSARs differed in their selection of molecular descriptors when describing the behaviors of chemical transversing across the viable epidermis. In the early days of \(Kp\) QSAR development, the molecular descriptors considered as significant factors influencing the transdermal transport of chemical and accordingly selected in the model were typically those of a measurable physicochemical property, e.g., the molecular weight (\(MW\)), melting point (\(MP\)), and logarithm of octanol-water partition coefficient (\(\log K_{OW}\)). The descriptors presented in the \(Kp\) QSARs have now shifted to emphasizing the processes involved in the arrangement of atomic occupation in molecular space (e.g., molecular volume, \(MV\)) and electronic distribution in that space (e.g., hydrogen bonding). This current study aimed to define a \(Kp\) QSAR using the state-of-the-art modeling technique to characterize, at a molecular level, the physical and chemical mechanisms involved in the transdermal permeation of chemical.

2. Research design and methods

To develop the \(Kp\) QSAR, the values of \(Kp\) for 158 structurally diverse chemicals of industrial and pharmaceutical application, as originally reported in Flynn [3] and Wilschut et al. [4], were selected and included in the QSAR training set. These \(Kp\) values were determined \textit{in vitro} using human cadaver skin. However, the exact procedures or experimental details in determining these \(Kp\)s were not reported in the original databases, and consequently inter-laboratory uncertainty and methodological variation among these \(Kp\)s were expected [5]. The \(MW\), \(\log K_{OW}\), two-dimensional chemical structure, and experimentally determined \(\log Kp\) were collected for each chemical: the \(MW\) and \(\log K_{OW}\) were extracted from the Syracuse Research Corporation PhysProp Database [6], the chemical structure extracted from the US National Library of Medicine TOXNET ChemIDplus Database (TOXNET) [7] or the National Institute of Standards and Technology’s Chemistry WebBook (NIST WebBook) [8] when the structure was unavailable from the TOXNET, and the human skin \(Kp\) extracted from Patel et al. [5]. Fig. 1 shows the distribution of the \(MW\) and \(\log Kp\) for the chemical included in the training set of \(Kp\) QSAR development; the \(MW\) ranged from 18.0 to 764.9 and the \(\log Kp\) from –6.1 to –0.19.

The development of QSAR typically included the following stages: the development of database for molecular structures of training compounds, the calculation and selection of molecular descriptors based on the structures, and the establishment of QSAR via statistical treatment such as the regression techniques. In this study, the molecular structures of the training compounds extracted from the TOXNET or NIST WebBook were graphically transformed and optimized in the HyperChem® Molecular Modeling System [9]. The molecular mechanics calculations involved in optimization were first performed using MM+ force field to rapidly optimize the molecular geometries with lower optimization accuracy, and then
the semi-empirical calculations were carried out using the routines AM1 to complete the full geometry optimization and improve the optimization accuracy. In the next step the Dragon® software [10] was used to calculate the molecular descriptors for all chemical compounds according to their optimized chemical structure. The Dragon® software in Version 5.5 calculated up to 3,224 descriptors for every molecule. The molecular descriptors yielding the same numerical value for all the training compounds were considered of low statistical relevance and hence removed from further test. With the removal of irrelevant descriptors, a total of 1,556 molecular descriptors remained in the test. These descriptors served as candidates for the regressor variables in a multiple linear regression (MLR) model. The MLR model was built in the statistical software MATLAB® [11] from a potentially large number of regressors, in this case the molecular descriptors, to indicate all the regressors considered essential in quantitatively describing the dependent variable, i.e., the target biological effect $K_p$. The molecular descriptors of key significance from the initial MLR model were recognized by the stepwise regression of log $K_p$ against each of the 1,556 descriptors and then by the selection of the descriptors showing a correlation coefficient greater than 0.1. In the final step, various combinations of descriptors qualified in the correlation tests were attempted by the MLR in iterated regression to identify the combination(s) that presented the optimal coefficient of determination ($R^2$) in the regression and best predicted the log $K_p$.

3. Results and discussion

Equation (1) shows the stepwisely regressed MLR model that described the $K_p$ QSAR established in this study, consisting of four molecular descriptors. These descriptors were statistically significant to the transdermal permeation behaviors of chemical and exerted their influence via: (1) electrostatic interactions between electric quadrupoles of van der Waals forces (descriptor name in Dragon®: QXXp); (2) partitioning of solutes between lipophilic vs. hydrophilic phases (ALOGP); (3) antineoplastic-like property at 80% similarity (Neoplastic-80); and (4) frequency of carbon-nitrogen bond at a topological distance of 06 (F06[C-N]) (Table 1). This four-descriptor MLR model fit the observed $K_p$ data with an $R^2$ of 0.828; the mean percentage error was 18.8%, a level comparable to those observed in the experimental data. Fig. 2 demonstrated the relative distribution of the predicted $K_p$ values against their experimentally observed counterparts for all of the 158 compounds in the training dataset.

$$
\log K_p = -3.0479 (\pm 0.0544) - 0.0065 (\pm 0.0006) \text{QXXp} + 0.6459 (\pm 0.0283) \text{ALOGP} \\
- 1.7541 (\pm 0.1115) \text{Neoplastic-80} + 0.2169 (\pm 0.0407) \text{F06[C-N]}
$$

(1)

Table 1. The Dragon® name, classification, and definition of molecular descriptors included in the quantitative structure-activity relationship established for prediction of skin permeation coefficient

| Molecular Descriptor | Type               | Definition                                              |
|----------------------|--------------------|---------------------------------------------------------|
| 1 QXXp               | Geometrical descriptors | Qxx COMMA2 value/weighted by atomic polarizabilities    |
| 2 ALOGP              | Molecular properties | Ghose-Crippen octanol-water partition coefficient       |
| 3 Neoplastic-80      | Molecular properties | Ghose-Viswanadhan-Wendoloski antineoplastic-like index at 80% |
| 4 F06[C-N]           | 2D frequency fingerprints | frequency of C - N at topological distance 06          |

In the $K_p$ QSARs summarized in Lian et al. [2], a linear correlation was frequently observed between the skin permeability and the lipophilicity/molecular size of the solute. As a general trend, the skin $K_p$ QSARs established using the $K_p$ data reported in Flynn [3] frequently considered a combination of the physicochemical descriptors $K_{OW}$ and $MW/MV$ as being mechanistically relevant and adequate in interpreting transdermal transport of chemical molecules [12]. In our study, the inclusion of the molecular descriptor Ghose-Crippen octanol-water partition coefficient in the QSAR shown in Equation (1) also suggested that this $K_p$ QSAR favored the lipophilicity as a key molecular property influencing the cutaneous permeation of chemical. As shown in Fig. 3, the log $K_p$ of the training compounds was considered in the stepwise regression as moderately correlated to the Ghose-Crippen octanol-water partition coefficient ($0.3 < r < 0.7$). However, for the other molecular property typically favored by a $K_p$ QSAR, the MW, in this study its correlation with the observed log $K_p$ was of a level insufficient to be statistically considered as correlated ($|r| < 0.3$).
Fig. 2. Distribution of predicted logarithmic skin permeation coefficient (log $K_p$) against experimentally observed log $K_p$ for 158 chemicals employed in developing quantitative structure-activity relationship.

Fig. 3. Distribution of experimentally observed logarithmic skin permeation coefficient (observed log $K_p$) against molecular descriptors (a) Ghose-Crippen octanol-water partition coefficient (ALOGP) and (b) molecular weight ($MW$).

4. Conclusions

The health hazard arising from skin exposure to toxic industrial chemical was typically assessed by determining the $K_p$ of chemical in the stratum corneum. However, limited by constraints in the experimental techniques this evaluation had not been sufficiently performed. This study developed a $K_p$ QSAR based on 158 chemical substances of known $K_p$ in vitro using human skin. The four descriptors in the current QSAR described the electrostatic interactions between electric quadrupoles of van der Waals forces, the chemical’s lipophilicity and antineoplastic property, and the frequency of carbon-nitrogen bonding at a fixed topological distance. The current QSAR considered log $K_{OW}$ a direct influence on the transdermal permeation of chemical. These findings suggest a capacity of the QSAR model reported in the current study to
serve as an alternative source of $K_p$ information and as a potential tool of dermal hazard characterization, particularly when experimentally determined $K_p$ are not readily available.

References

[1] Department of Health and Human Services (DHHS), 2000. Objective for Improving Health Focus Area 20-8: Reduce Occupational Skin Diseases or Disorders among Full-Time Workers. Healthy People 2010, Vol. II, 2nd Ed. DHHS, Washington, DC, USA.

[2] Lian, G., Chen, L., Han, L., 2008. An Evaluation of Mathematical Models for Predicting Skin Permeability, Journal of Pharmaceutical Sciences 97, p. 584.

[3] Flynn, L., 1990. Physicochemical Determinants of Skin Absorption, in “Principles of Route-to-Route Extrapolation for Risk Assessment” T. Gerrity, C. Henry, Editors. Elsevier Science Publishing Co., New York, p. 93.

[4] Wilschut, A., ten Berge, W., Robinson, P., McKone, T., 1995. Estimating Skin Permeation. The Validation of Five Mathematical Skin Permeation Models, Chemosphere 30, p. 1275.

[5] Patel, H., ten Berge, W., Cronin, M., 2002. Quantitative Structure–Activity Relationships (QSARs) for the Prediction of Skin Permeation of Exogenous Chemicals, Chemosphere 48, p. 603.

[6] Syracuse Research Corporation (SRC), 2010. PhysProp Database, available from http://www.syrres.com/esc/physprop.htm. SRC, Arlington, VA, USA.

[7] US National Library of Medicine (NLM), 2011. Toxicology Data Network (TOXNET) ChemIDplus Database, available from http://toxnet.nlm.nih.gov/. NLM, Bethesda, MD, USA.

[8] National Institute of Standards and Technology (NIST), 2011. Chemistry WebBook, available from http://webbook.nist.gov/chemistry/. Gaithersburg, NIST, MD, USA.

[9] Hypercube Inc., 2008. Reference Manual. HyperChem® Release 8.0 for Windows® Molecular Modeling System. Hypercube Inc., Waterloo, ON, Canada.

[10] Talete srl., 2007. DRAGON User Manual. DRAGON for Windows and Linux 2007, Version 5.5, available from http://www.talete.mi.it/. Talete srl, Milano, Italy.

[11] The MathWorks Inc., 2009. MATLAB Statistical Toolbox v7.7 – User Guide. The MathWorks, Inc., Natick, MA, USA.

[12] Fitzpatrick, D., Corish, J., Hayes, B., 2004. Modelling Skin Permeability in Risk Assessment–the Future, Chemosphere 55, p. 1309.