RESEARCH ARTICLE

Evaluation of selected quantitative structure permeability relationship (QSPR) based mathematical models for the prediction of skin permeability of *Camellia sinensis* (tea) compounds

Sudeshna Kundu[a], Anusua Singh[a,b], Sharmistha Samanta[a], Dipankar Chattopadhyay[c], Raghwendra Mishra[b*]
Roshnara Mishra[a*]

ABSTRACT

Background: Skin permeability coefficient (logK_p) is a major determinant for topical drugs. *In-vitro* and *ex-vivo* determination of logK_p is expensive, time and labor-intensive, and difficult to apply to large databases. QSPR models derived from the statistical correlation between descriptors of the compounds with the *in-vitro* or *ex-vivo* permeation data are used extensively. The vast number of QSPR equations makes the selection of a particular equation to screen a database difficult. *Objective:* This study has evaluated common descriptor-based equations to select the best suitable equation for screening the phytochemical library of *Camellia sinensis*. *Methods:* Seven QSPR-based models were used to estimate and compare the logK_p of tea compounds. The best method was selected with respect to the gold standard logK_p. *Result:* The model of Potts and Guy showed close proximation with the gold standard logK_p and had the highest association along with least RMSE value and least deviation. According to this method, approximately 37.38, 35.35 and 27.27% of the tea compounds were found to have high, good and poor logK_p, respectively. *Conclusion:* Potts and Guy equation can be effectively used to screen the phytochemical library of *Camellia sinensis*. This study has potential applications in the field of topical medicine and cosmetics.

Keywords: Skin permeation coefficient, QSPR, topical, *Camellia sinensis,* tea, phytochemical.

Indian Journal of Physiology and Allied Sciences (2023); DOI: 10.55184/ijpas.v75i03.143 **ISSN:** 0367-8350 (Print)

INTRODUCTION

Drug delivery through the skin offers an attractive and alternative route of drug administration over oral and parenteral drug delivery. It bypasses the first-pass metabolism and overcomes the limitations of oral drugs like, GI degradation, hepatic clearance etc. Moreover, this non-invasive route is convenient and preferred over the parenteral route of administration.[1] Unlike systemic application, local administration of topical drug maximizes therapeutic efficacy by increasing local tissue concentration and minimizing adverse effects of nonspecific targeting.[2] In spite of such advantages, poor skin permeation of drug candidates is a major limitation for topical route. The stratum corneum, the thickest layer containing numerous coverings of keratinized corneocytes, is the primary barrier for drug permeation through the skin.[3] However, some provisions exist for transferring natural compounds across the skin, including the intercellular, intracellular and follicular pathways. The intercellular path facilitates the transmission of hydrophilic drugs, whereas the intracellular path is suitable for the transport of lipophilic drugs. The follicular or trans-appendageal path allows the rapid transfer of drugs directly to the infundibulum region.[4]

The barrier function of skin, imparted by the unique arrangement of hydrophilic keratin filaments compactly packed with hydrophobic lamellar lipids,[5] presents a challenge to the study of skin permeation. Different types of skin permeation measurements have been developed to assess the dermal absorbed dose. Among them the most common are- percent absorbed and permeability coefficient (logK_p). Measurement of the former requires careful surface area control and is highly dependent on the magnitude and duration of exposure. Therefore, being independent of time, volume, exposure concentration and cross-laboratory comparison, logK_p is increasingly used.[6]

LogK_p, the major determinant of the bioavailability of topical drug candidates, is derived from Fick's law and is determined

---

[a]Department of Physiology, University of Calcutta, Kolkata, West Bengal, India

[b]Department of Physiology, Ananda Mohan College, Kolkata, West Bengal, India

[c]Department of Polymer Science and Technology, University of Calcutta, Kolkata, West Bengal, India

***Corresponding author:** Roshnara Mishra, Department of Physiology, University of Calcutta, Kolkata, West Bengal, India, Email: roshnaramishra@gmail.com

How to cite this article: Kundu S, Singh A, Samanta S, Chattopadhyay D, Raghwendra Mishra R, Mishra R. Evaluation of selected quantitative structure permeability relationship (QSPR) based mathematical models for the prediction of skin permeability of *Camellia sinensis* (tea) compounds. *Indian J Physiol Allied Sci*. 2023;75(3):20-27.

Conflict of interest: None

Submitted: 22/05/2023 **Accepted:** 09/06/2023 **Published:** 25/09/2023

by several *in-vitro* and *ex-vivo* methods using polymer membranes or excised mammalian skin. Besides ethical considerations, these methods are expensive and time and labor-intensive. While for individual drug candidates, these methods are still worth the trouble; these are, therefore, difficult to apply for large databases. Quantitative structure permeability relationship (QSPR) based mathematical models for $\log K_p$ determinations are used extensively to screen compound libraries. The output of the QSPR model is in form of an equation constructed from the statistical correlation between physicochemical or molecular descriptors of the compounds with the *in-vitro* or *ex-vivo* permeation data.

In the last few decades, many QSPR based mathematical models for predicting skin permeability with different degrees of accuracy and limitations have been reported and validated. These models have used different combinations of mechanistic and empirical descriptors, *in-vitro* or *ex-vivo* permeation data and different statistical or machine learning tools for correlation study. However, every model has its limitations. No model can be used universally as the complexity of skin permeation physiology is too great to be bound by the assumptions on which a model is built. This presents a unique challenge for application of QSPR models in large databases.

This study has evaluated seven QSPR based models that have used common descriptors and common assumptions and are used extensively by in silico tools to select the equation best suitable for screening the phytochemical library of *Camellia sinensis*.

Though more famous as a beverage, tea has been used topically since ancient times. Traditional tea baths were famous in ancient China for skin and hair and were also used in detoxification. *Camellia sinensis* and its compounds are used as topically applied cosmetics and have been reviewed in detail.[7] Black tea dressing has been reported to be beneficial in facial dermatitis.[8] Topical formulations of green and black tea have been reported to be effective against skin cancer,[9,10] UV-induced damage,[11,12] wound healing,[13] as well as have anti-aging effects.[14,15] The antibacterial effect of caffeine has been clinically tested in psoriasis.[16] The benefits of tea, not only as a beverage but also as a topical remedy and cosmetic, warrant assessment of the skin permeability profile of its compounds library.

The present study is designed to evaluate tea compounds' skin permeability by using well-documented mathematical models based on a common set of limited and easily obtainable parameters, compare them, and select the best suitable one based on the standard methodology.

MATERIALS AND METHODS

Construction of Phytochemical library of *Camellia sinensis* and Preparation of Ligands for Qikprop Analysis

Tea phytochemicals were enlisted after extensive literature searching. Beside this, databases such as IMPPAT (https://cb.imsc.res.in/imppat/), Dr. Duke's Phytochemical and Ethnobotanical database (https://phytochem.nal.usda.gov/) and PubChem (https://pubchem.ncbi.nlm.nih.gov/) were also utilized to screen out a phytochemical library of 693 compounds (Figure 1). Their SDF structures were downloaded from PubChem and subsequently prepared using the LigPrep module of Schrödinger suite version 2020-3, considering all possible stereochemical, ionization and tautomeric variations using the OPLS3e force field. All the possible ionization and tautomeric states between pH 6.8 and 7.2 were generated using Epik module and the optimized ligands were used for QikProp studies.[17]

Prediction of Skin Permeability using Different Mathematical Models

Skin permeability coefficient ($\log K_p$) was assessed to investigate cutaneous absorption of tea compounds. The QSPR equations for prediction of $\log K_p$ (Table 1) which use either or both octanol-water partition coefficient ($\log K_{ow}$) and molar mass (MW), two easily obtainable descriptors, were selected for this study. Log transformation of predicted K_p values of Mitragotri and Hatanaka *et al.* equation was carried out for common scaling of data. The unit conversion, from cm/hr to cm/sec, was carried out in case of Vecchia and Bunge, QikProp and DERMWIN^{TM} predicted $\log K_p$ values. The predicted skin permeation coefficient ($\log K_p$) using different mathematical models and skin permeation coefficient ($QPlogK_p$) yielded by QikProp (referred as QikProp predicted $\log K_p$ value hereafter) were used for further analysis.[17]

For Mitragotri's equation, the second factor molecular radius (r), was calculated using MW, as mentioned by Lian *et al.* (22) as, $\frac{4}{3}\pi r^3 = 0.91 \times MW$.

Selection of Mathematical equation for Prediction of $\log k_p$

Due to the lack of any agreeable gold standard method for the determination of $\log K_p$ value, and a pairwise analysis of reliability and agreement between the selected methods

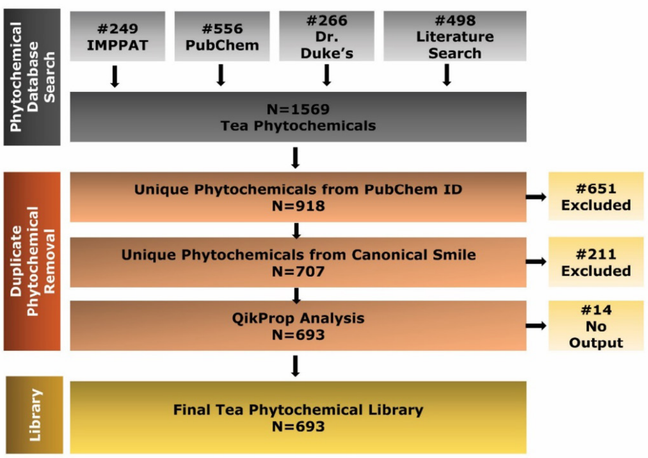

Figure 1: Schematic representation of construction of *Camellia sinensis* phytochemical library

Table 1: Selected mathematical QSPR models and their equations

QSPR based models	*Equations*
Hatanaka *et al.*[18]	$K_p = 4.78 \times 10^{-7} \times K_{OW}{}^{0.589} + 8.33 \times 10^{-8}$
Potts and Guy[19]	$logK_p = -6.3 + 0.71 \times logK_{OW} - 0.0061 \times MW$
Mitragotri[20]	$K_p = 5.6 \times 10^{-6} \times K_{OW}^{0.7} e^{(-0.46r^2)}$
Vecchia and Bunge[21]	$\log K_p = -2.44 + 0.514 \times logK_{OW} - 0.005 \times MW$
Lian *et al.*[22]	$logK_p = -5.2 + 0.7 \times logK_{OW} - 0.072 \times MW^{2/3}$
DERMWIN^{TM}[23]	$logK_p = -2.8 + 0.66 \times logK_{OW} - 0.0056 \times MW$

Where, Kow: octanol/water partition coefficient; MW: the molecular weight of the compound; r: molecular radius of the compound. DERMWIN is a module for estimation of skin permeability coefficient in the EPI Suite (https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface)

will result in a cumbersome output (21 pairs of different statistical analyses) which will be hard to interpret, thus, there is a need of a consensus standard for comparison. Also, the experimental methods do suffer from inherent limitations (in this case, pH, temperature, skin type etc.) and as there is a lack of experimental data of such a large chemical set, we have to look for other feasible options for better understanding and interpretation of the results. The average of the predicted $logK_p$ values of the selected equations (referred to as, average $logK_p$ hereafter) is said to be consensus representative in such cases and used across different fields in biology,[24-29] and references there in). The average $logK_p$ was calculated and used as the gold standard in the present study. The predicted $logK_p$ values were then compared with the average $logK_p$ value to estimate agreement and selection of best suitable model.

Skin Permeability Profile of *Camellia sinensis* Compounds

The best equation selected based on the performance and agreement with the average $logK_p$ values, was then employed to evaluate the skin permeation profile of tea compounds on a scale constructed with respect to the known standard mammalian skin permeable phytochemicals- naringenin, an aglycone flavonoid, and its corresponding glycoside naringin according to Chuang *et al.*[30]

Statistical Analysis

The predicted $logK_p$ values were summarized as mean, standard deviation and 95% confidence interval of mean (95% CI). The variation in the predicted $logK_p$ value was assessed using analysis of variance (ANOVA) followed by appropriate *post hoc* analysis depending on whether the data met the assumptions. The association between the predicted $logK_p$ values was assessed using the correlation coefficient. The accuracy of the prediction of the gold standard $logK_p$ value was estimated using the coefficient of determination (R^2) and root mean square error (RMSE), a residual statistic, which gives a good idea of both bias and spread of the data. Finally, the Bland and Altman plot analysis was conducted and the 95% limit of agreement (mean ± 1.96 × standard deviation ranges of the respective differences) between predicted $logK_p$ values compared to the gold standard value was evaluated.[28,31] The coefficient of variation (%CV), a measure of relative dispersion, of the difference between the QSPR predicted $logK_p$ values and standard $logK_p$ value was also estimated. The reliability of the predicted $logK_p$ values was estimated using the intraclass correlation coefficient (ICC) (2-way mixed-effect model), in term of both consistency and absolute agreement.[32] A two-tail $p<0.05$ was considered as statistically significant. All the analyses were carried out using MS Excel (version 2019) and statistical program packages OriginPro 2021b, OriginLab Corporation, Northampton, MA, USA.

RESULT

The prepared phytochemical library of *Camellia sinensis* comprises 693 compounds belonging to different chemical classes, with molecular weight ranging between 31 to 1322 Da and $logK_{ow}$ between -7 to +18. The tea phytochemicals library, classified as per Cumming and Rucker., 2017[33] was found to be consisting of 41 (5.92%) extremely hydrophilic phytochemicals ($logK_{ow} \leq -3$), 13 (1.88%) extremely hydrophobic phytochemicals ($logK_{ow} \geq +10$) and remaining 639 (92.2%) compounds with $logK_{ow}$ values ranging between -3 to +10.

Using the selected equations, the logKp values were then calculated from the QikProp generated molecular weight and logKow. The distribution of the predicted $logK_p$ values of the tea compounds obtained from different mathematical equations, QikProp and the average $logK_p$ were depicted in Figure 2. Except for the equations given by Hatanaka *et al.* and DERMWIN^{TM}, all the predicted $logK_p$ values showed similar distribution patterns.

The mean, standard deviation and 95% confidence interval of the mean of predicted $logK_p$ were tabulated in Table 2. The average $logK_p$ value was calculated as the arithmetic mean of the predicted $logK_p$ values of the seven QSPR methods and used as the gold standard for further comparison. As the

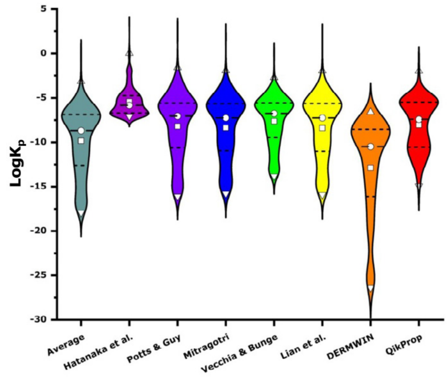

Figure 2: Violin plot depicting the distribution of predicted $logK_p$ values of selected mathematical equations, QikProp and the average $logK_p$. €- Mean, o- Median, Δ- 99th percentile, ∇-1st percentile.

Table 2: Predicted logK_p values of the *Camellia sinensis* phytochemical library from different equations and average logK_p

LogK_p	*Mean*	*Standard Deviation*	*95% Confidence Interval of mean*		*p-value (Welch ANOVA)*
			Lower Bound	*Upper Bound*	
Average	-9.83	3.831	-10.111	-9.539	<0.001
Hatanaka *et al.*[18]	-5.42	1.651	-5.545	-5.299	
Potts and Guy[19]	-8.20	3.686	-8.475	-7.925	
Mitragotri[20]	-8.37	3.604	-8.643	-8.105	
Vecchia and Bunge[21]	-7.63	2.841	-7.843	-7.419	
Lian *et al.*[22]	-8.40	3.646	-8.668	-8.124	
DERMWIN™	-12.877	5.592	-13.294	-12.460	
QikProp	-8.05	2.979	-8.271	-7.827	

assumption of equivalence of variance was violated (Levene statistics, $p<0.001$) Welch ANOVA followed by Games-Howell *post hoc* pairwise comparison was conducted to evaluate any significant variation of predicted logK_p values between the studied methods. Predicted logK_p values followed the order- Hatanaka *et al.*>Vecchia and Bunge>QikProp>Pots and Guy>Mitragotri≈Lian *et al.*>DERMWIN™ and varied significantly between the selected methods (Welch ANOVA, $p<0.001$). On pairwise comparison, insignificant variation was observed among the predicted logK_p values of Potts and Guy, Mitragotri, Lian *et al.* and QikProp ($p= 0.552$-1.000). The predicted logK_p value obtained in case of Hatanaka *et al.* equation was found to be significantly higher compared to all other methods ($p<0.001$) followed by the equation given by Vecchia and Bunge ($p=0.132$-<0.001), whereas DERMWIN™ yields significantly lower logK_p compared to all other methods ($p<0.001$).

The scatter plot analysis of predicted logK_p values for different pair of equations (not shown) revealed that they have monotonic association, i.e. the data points are moving in the same relative direction but not with a constant rate (as in case of linear relation). Hence Spearman rho (ρ) correlation analysis was conducted to evaluate the association between the predicted logK_p values obtained from different equations and the selected gold standard logK_p value (Figure 3). All predicted logK_p values showed significant positive association (Spearman $\rho\geq0.658$, $p<0.001$) with one another. Potts and Guy, Mitragotri, Lian *et al.* and Vecchia and Bunge predicted logK_p values had ρ values of 0.996 and above with each other and the consensus average logK_p value, whereas DERMWIN™ predicted logK_p value showed comparatively lower association ($\rho\leq0.878$, $p<0.001$) with all others. The consensus gold standard selected for this study, i.e. the average logK_P value showed strong correlation with all the predicted logK_p values ($\rho=0.878$-0.998, $p<0.001$).

From the R^2 values for standard logK_p, as predicted from the QSPR equations, it was evident that the average logK_p value could be explained well by all the predictors (Table 3), which was in accordance to the correlation data. The lowest RMSE value for predicted average logK_p was observed from the equation of Potts and Guy (RMSE= 0.005) followed by Lian *et al.* (RMSE=0.007), whereas Hatanaka *et al.* equation showed highest RMSE value (RMSE= 0.125).

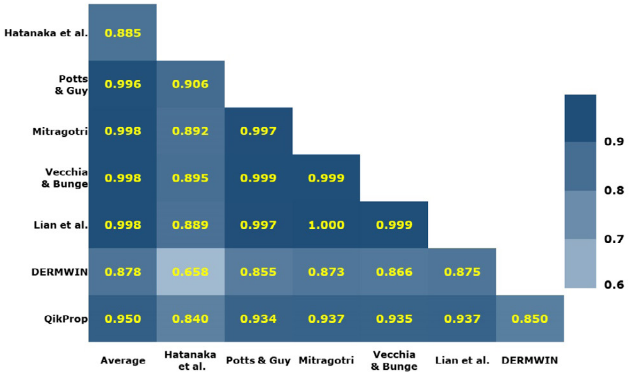

Figure 3: Correlogram depicting the Spearman ρ values among the predicted logK_p from different models and the gold standard logK_p (average logK_p) values.

The result of correlation and regression analysis indicated that there was a significant association between the criterion (standard logK_p values) and predictors (the predicted logK_p values from selected equations) but failed to provide any information regarding the limit of agreement and reliability of the data compared to the gold standard method. Bland and Altman plot analysis was conducted in order to assert the agreement between the predicted logK_p values from the gold standard one.[31] The consensus gold standard average logK_p value showed significant difference compared to all the methods used for the prediction of logK_p values (Games-Howell *post hoc* analysis, $p<0.001$). The coefficient of variation (%CV), a measure of dispersion of dataset, of the difference between predicted and average logK_p values (represented on the y-axis of Figure 4) showed appreciable variation (21.15–90.59%) across the QSPR methods. The least dispersion was noted in case of Lian *et al.* (%CV=21.15%) closely followed by Potts and Guy (%CV= 21.31%). Despite the varied dispersion across the data set, the predicted logK_p values showed good agreement with average logK_p as indicated by the number of compounds falling between 95% limit of agreement, the predicted logK_p and standard average logK_p values (Table 4) ranging 86.75–97.98%. The predicted logK_p values and standard logK_p values showed excellent consistency

Table 3: Accuracy of standard logK_p value obtained from each formula compared to the selected gold standard logK_p value

QSPR Equations	*Average logK$_p$ (n=693)*	
	R^2	*RMSE*
Hatanaka *et al.* (18)	0.5624	0.125
Potts and Guy (19)	0.9930	0.005
Mitragotri (20)	0.9953	0.008
Vecchia and Bunge (21)	0.9954	0.011
Lian *et al.* (22)	0.9954	0.007
DERMWIN^{TM}	0.8887	0.080
QikProp	0.8436	0.010

(ICC=0.975, mean measurement: K=8, consistency, 2-way mixed-effect model) and excellent agreement (ICC=0.929, mean measurement: K=8, absolute agreement, 2-way mixed-effect model) and was in accordance with the above results (21). Except for Hatanaka predicted logK_p values (ICC=0.410), all other QSPR equation predicted logK_p values showed good to excellent reliability (0.845≤ICC≤0.963), mean measurement: K=2, absolute agreement, 2-way mixed-effect model) compared to the consensus standard (Table 4).

The intraclass correlation coefficient was calculated based on mean measurement (K=2) of the used QSPR equation and the consensus standard, for absolute agreement using a 2-way random mixed-effect model.

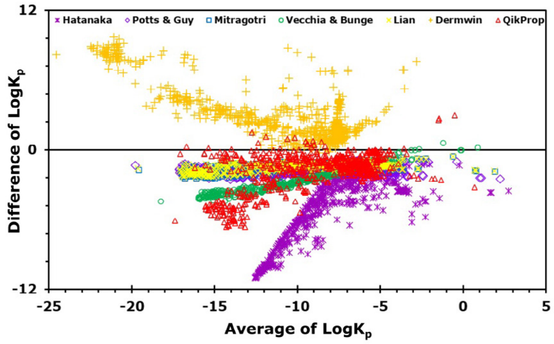

Figure 4: Bland-Altman agreement plot for difference between predicted logK_p values obtained from different equations and respective gold standard data- average logK_p.

DISCUSSION

Skin, the body's largest organ, forms a unique and flexible interface between the body's internal milieu and the external environment; as a potential barrier, skin protects the body from foreign compounds. Stratum corneum, the outermost permeability barrier of the skin, is made of multilayers of hydrophilic keratin filaments embedded in a hydrophobic lamellar lipid matrix. The type and the amount of lipid in the stratum corneum depends on the site of the body and, it is generally accepted that skin permeability is affected by this lipid layer.[34,35] Michaels *et al.*, 1975[36] showed that several drugs had significant skin permeability and determined their stratum corneum diffusion coefficients. The main limiting factor for this process is the slow diffusion through the dead layer of stratum corneum. Several investigators have used the published human stratum corneum permeability coefficient (K_p, often expressed as logK_p) data to predict the skin permeability and examined the effect of the structural parameters of penetrants on the permeability[19,37] which led to the development of QSPR models. QSPRs are useful in predicting the behavior of novel compounds and provide insights into mechanisms of activity. A current trend in QSPR studies is the use of theoretical molecular descriptors that can be calculated directly from molecular structure.

The descriptors used in the QSPR development are mostly measurable and easily obtainable physicochemical properties like- molecular weight (MW), melting point (MP), and logK_{ow}. The MW and logK_{ow} are often the key- and in most cases the only- descriptors in the correlation-based QSPRs, such as the Potts and Guy method (38). Though using descriptors on the basis of ease of measurement undermines important influencers, these methods prevail in practice. Most *in silico* tools (such as- DERMWIN, QikProp, SwissADME etc.) use Potts and Guy equation or equation based on the same dataset used by them.[23,24,39]

This study chose six different QSPRs consisting of only MW and logKow as descriptors, along with QikProp generated logKp values, to evaluate the most suitable QSPR to predict the skin permeability. The profile of selected seven QSPR based models is given in Table 5. Hatanaka *et al.* provided an equation for predicting drugs' steady state permeation

Table 4: Profile of logK_p of tea phytochemicals within 95% limit of agreement range (Bland-Altman Plot) and Intraclass correlation coefficient (ICC) estimate of studied QSPR models

Equations	%CV	Difference of LogK_p(n=693)		Number (Percentage) of Tea phytochemicals falling in the 95% limit of agreement
		Mean – 1.96 SD	Mean + 1.96 SD	
Hatanaka *et al.* (18)	63.90	-9.92	1.11	637 (91.92%)
Potts and Guy (19)	21.31	-2.30	-0.95	669 (96.54%)
Mitragotri (20)	23.50	-2.12	-0.78	633 (91.34%)
Vecchia and Bunge (21)	46.26	-4.18	-0.21	679 (97.98%)
Lian *et al.* (22)	21.15	-2.02	-0.84	635 (91.63%)
DERMWIN^{TM}	77.22	-1.57	7.67	622 (86.75%)
QikProp	90.59	-4.93	1.38	635 (91.63 %)

rate based on their model of two parallel skin permeation pathways of lipid and pore. They had mentioned that the permeability coefficient is correlated to the partition coefficient and proposed an equation based on K_{ow}. Potts and Guy provided a mechanistically based model, preferable for the compounds ranging between MW 18-750Da and $\log K_{ow}$ -3 to +6. They considered the lipid matrix as the pathway of skin permeation. Based on the analytical solutions of diffusion Mitragotri proposed a mechanistic model. In this model, four pathways were taken for consideration- free-volume diffusion through lipid bilayers, lateral diffusion along lipid bilayers, diffusion through pores, and diffusion through shunts. Mitragotri's equation describes free volume diffusion of the lipophilic chemicals ($\log K_{ow} > 1$) which was found to be in perfect agreement with the Potts Guy method. Mitragotri also explained the limitation of Potts Guy QSPR mechanistically and mathematically by showing the permeation of hydrophilic solutes by diffusion through aqueous pores and giving a correction on Potts and Guy equation. Lian *et al.*, 2008 provided a modified form of Mitragotri's equation by substituting the solute radius of a molecule with the molecular weight for the diffusion in lipid bilayer and given the equation resemblance to the equation of Potts and Guy. Vecchia and Bunge established a model based on $\log K_{ow}$ and MW, providing a simple equation for reasonably estimating the stratum corneum permeability coefficient. They had presented diverse MW ranging from 18-584Da, and $\log K_{ow}$ ranging from -3.1 to +4.6 to develop the equation. Two software-based models, DERMWIN^{TM} and QikProp taken in this study, developed their QSPRs using the dataset of Potts and Guy but they did not provide enough information in their user guidelines.[40,41]

There is a great discrepancy between the predicted $\log K_p$ using different models as well as the superiority of a particular QSPR model basically due to their empirical nature and the experimental conditions under which data were collected.[22,42] Hence, while selecting a particular method, its agreement, reliability and reproducibility with some gold standard method, is warranted.[31,32] Simple association statistics, which are a measure of relationship but not the differences alone, fails to assess the comparability. In the present study, along with association statistics, Bland-Altman plot analysis was conducted for agreement and intraclass correlation coefficient, a reliability index, was estimated to select an appropriate QSPR method for the prediction of $\log K_p$ values having better consistency and agreement with the gold standard $\log K_p$ value.

Based on the overall performance of different statistical parameters of the predicted $\log K_p$ values in relation to the gold standard $\log K_p$ value, the QSPR equation proposed by Potts and Guy was selected as the method of choice for the prediction of $\log K_p$ value based on lowest RMSE value and %CV (dispersion of its difference from standard $\log K_p$ value) associated with better overlap with other predicted $\log K_p$ values, 95% limit of agreement, and comparable Spearman ρ, R^2, consistency and absolute agreement, for further use with the tea phytochemical library. The Potts and Guy equation predicted $\log K_p$ values of the tea phytochemicals

Table 5: Descriptive chart of QSPR based models taken in the study

Sl. No.	*Models*	*Reference range of compounds*	*Database used*	*R^2 value of the equation*	*Limitations*
1	Hatanaka *et al.* (18)	MW:130.08 to 375.85Da; $\log K_{ow}$: -4.7 to +4.4	*In-vitro* dataset of hairless rat skin and artificial membrane (n=17)	Not reported	Could not predict the permeation of the compounds which are hydrophilic and high molecular weighted
2	Potts and Guy (19)	MW: 18 to >750Da; $\log K_{ow}$: -3 to +6	*In-vitro* dataset of human epidermis (n~90) from Flynn., 1990 (37)	0.676	Under predicts the skin permeability of hydrophilic compounds
3	Mitragotri (20)	MW: 18Da to 150KDa; $\log K_{ow}$: -6.9 to +5.49	*In-vitro* dataset of mammalian epidermis (n=120) from Jhonson *et al.*, 1997 (41)	0.698	Under predicts the skin permeability of highly hydrophilic compounds (K_{ow}<0.01) and ignored hydrophilic pathway
4	Vecchia and Bunge (21)	MW:18 to 584Da; $\log K_{ow}$: -3.1 to +4.6	*In-vitro* dataset of human skin (n=127) gathered from multiple dataset	0.551	More permeability coefficient data needed to decide the mechanism of permeation of hydrophilic compounds
5	Lian *et al.* (22)	MW: 18 to 765Da; $\log K_{ow}$: -3.7 to +5.49	*In-vitro* dataset of human skin (n=124) gathered from multiple dataset	0.698	Modification of Mitragotri's method only
6	DERMWIN^{TM}	Not mentioned	Used data set from Potts and Guy., 1992 (19)	Not mentioned	Insufficient information (40)
7	QikProp	Not clearly mentioned	Used data set from Potts and Guy., 1992 (19)	0.78	Insufficient information

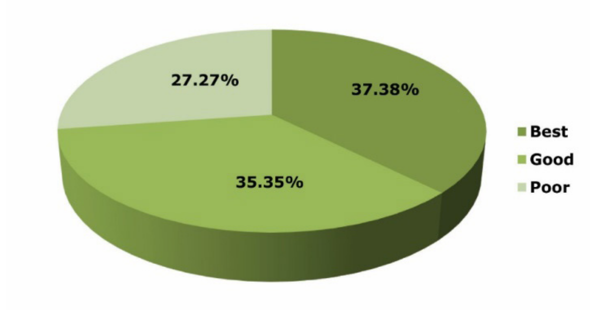

Figure 5: Pie-diagram depicting the percentage of tea compounds shown to have $\log K_p$ above naringenin (excellent), between naringenin and naringin (good) and below naringin (poor)

were then scaled on the basis of predicted $\log K_p$ values of the selected skin permeable standard phytochemicals (based on the report of Chuang *et al.*, 2017) (30)- naringenin ($\log K_p$=-6.17153; suggestive for high skin permeability boundary) and naringin ($\log K_p$=-10.1535; suggestive for poor permeability boundary) and their percutaneous administration profile were constructed (Figure 5). It was observed that, out of 693, 259 (37.38%) tea phytochemicals have a skin permeation rate higher than that of naringenin, whereas 245 (35.35%) components were found to have their skin permeability between the range of naringenin and naringin. As more than two-thirds of the tea phytochemicals possess appreciable skin permeability, as per *in silico* $\log K_p$ predicted value, it can be considered as a potent candidate for topical application and formulation.

Agreement and reliability facilitated the selection of one particularly successful model to predict $\log K_p$ using the mathematical equation provided by Potts and Guy. This model transforms the easily obtainable descriptors like, molar mass and partition coefficient of a known dataset as a resource of information to a more beneficial model to replace permeation testing for a wide range of compounds with an unknown dataset. Using this model, our prepared library of tea phytochemicals was found to have 72.73% skin-permeable compounds. For the rest of the compounds with poor permeability, carriers should be used for their successful topical and transdermal application. This study is limited on the small set of QSPR based models and their validation with the data obtained from the computational study. Therefore, further study is warranted along with the experimental data set for a better understanding of these models' performance.

REFERENCES

1. Alexander A, Dwivedi S, Ajazuddin, Giri TK, Saraf S, Saraf S, Tripathi DK. Approaches for breaking the barriers of drug permeation through transdermal drug delivery. *J Control Release*. 2012;164(1):26-40. DOI: 10.1016/j.jconrel.2012.09.017.
2. Leppert W, Malec-Milewska M, Zajaczkowska R, Wordliczek J. Transdermal and Topical Drug Administration in the Treatment of Pain. *Molecules*. 2018;23(3):681. DOI: 10.3390/molecules23030681.
3. Alkilani AZ, McCrudden MT, Donnelly RF. Transdermal Drug Delivery: Innovative Pharmaceutical Developments Based on Disruption of the Barrier Properties of the stratum corneum. *Pharmaceutics*. 2015;7(4):438-70. DOI: 10.3390/pharmaceutics7040438.
4. Iliopoulos F, Sil BC, Evans CL. The role of excipients in promoting topical and transdermal delivery: Current limitations and future perspectives. *Front Drug Deliv*. 2022;2(104984):1-10. DOI: 10.3389/fddev.2022.1049848.
5. Phatale V, Vaiphei KK, Jha S, Patil D, Agrawal M, Alexander A. Overcoming skin barriers through advanced transdermal drug delivery approaches. *J Control Release*. 2022;351:361-80. DOI: 10.1016/j.jconrel.2022.09.025.
6. Poet TS, McDougal JN. Skin absorption and human risk assessment. *Chem Biol Interact*. 2002;140(1):19-34. DOI: 10.1016/s0009-2797(02)00013-3.
7. Koch W, Zagórska J, Marzec Z, Kukula-Koch W. Applications of Tea (*Camellia sinensis*) and its Active Constituents in Cosmetics. *Molecules*. 2019;24(23):4277. DOI: 10.3390/molecules24234277.
8. Witte M, Krause L, Zillikens D, Shimanovich I. Black tea dressings - a rapidly effective treatment for facial dermatitis. *J Dermatol Treat*. 2019;30(8):785-9. DOI: 10.1080/09546634.2019.1573306.
9. Katiyar SK. Green tea prevents non-melanoma skin cancer by enhancing DNA repair. *Arch Biochem Biophys*. 2011;508(2):152-8. DOI: 10.1016/j.abb.2010.11.015.
10. Mittal A, Piyathilake C, Hara Y, Katiyar SK. Exceptionally high protection of photocarcinogenesis by topical application of (--)-epigallocatechin-3-gallate in hydrophilic cream in SKH-1 hairless mouse model: relationship to inhibition of UVB-induced global DNA hypomethylation. *Neoplasia*. 2003;5(6):555-65. DOI: 10.1016/s1476-5586(03)80039-8.
11. Seleem M, Abulfadl YS, Hoffy N, Lotfy NM, Ewida HA. Promising role of topical caffeine mesoporous gel in collagen resynthesis and UV protection through proline assessment. *Future J Pharm Sci*. 2022;8(1):27. DOI: 10.1186/s43094-022-00417-5.
12. Sopyan I, Permata RD, Gozali D, Syah ISK. Formulation of lotion from black tea extract (*Camellia sinensis* Linnaeus) as sunscreen. *Int J Pharm*. 2019;11(1):205-9. DoI: 10.22159/ijap.2019v11i1.29564.
13. Kouhihabibidehkordi G, Kheiri S, Karimi I, Taheri F, Bijad E, Bahadoram M, *et al*. Effect of white tea (*Camellia sinensis*) extract on skin wound healing process in rats. *World J Plast Surg*. 2021;10(1):85-95. DOI: 10.29252/wjps.10.1.85.
14. Lee KO, Kim SN, Kim YC. Anti-wrinkle Effects of Water Extracts of Teas in Hairless Mouse. *Toxicol Res*. 2014;30(4):283-9. DOI: 10.5487/TR.2014.30.4.283.
15. Chiu AE, Chan JL, Kern DG, Kohler S, Rehmus WE, Kimball AB. Double-blinded, placebo-controlled trial of green tea extracts in the clinical and histologic appearance of photoaging skin. *Dermatol Surg*. 2005;31(7 Pt 2):855-60. DOI: 10.1111/j.1524-4725.2005.31731.
16. Vali A, Asilian A, Khalesi E, Khoddami L, Shahtalebi M, Mohammady M. Evaluation of the efficacy of topical caffeine in the treatment of psoriasis vulgaris. *J Dermatolog Treat*. 2005;16(4):234-7. DOI: 10.1080/09546630510011801.
17. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des*. 2013;27(3):221-34. DOI: 10.1007/s10822-013-9644-8.
18. Hatanaka T, Inuma M, Sugibayashi K, Morimoto Y. Prediction of skin permeability of drugs. I. Comparison with artificial membrane. *Chem Pharm Bull (Tokyo)*. 1990;38(12):3452-9. DOI: 10.1248/cpb.38.3452.
19. Potts RO, Guy RH. Predicting skin permeability. *Pharm Res*.

1992;9(5):663-9. DOI: 10.1023/a:1015810312465.

20. Mitragotri S. Modeling skin permeability to hydrophilic and hydrophobic solutes based on four permeation pathways. *J Control Release*. 2003;86(1):69-92. DOI: 10.1016/s0168-3659(02)00321-8.
21. Vecchia BE, Bunge AL. Skin absorption databases and predictive equations. 2nd ed. New York: Marcel Dekker; 2003. 57–141.
22. Lian G, Chen L, Han L. An evaluation of mathematical models for predicting skin permeability. *J Pharm Sci*. 2008;97(1):584-98. DOI: 10.1002/jps.21074.
23. US-EPA. Estimation Programs Interface SuiteTM for Microsoft® Windows. Washington, 2018.
24. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*. 2017;7(1):42717. DOI: 10.1038/srep42717.
25. Orimadegun A, Omisanjo A. Evaluation of five formulae for estimating body surface area of nigerian children. *Ann Med Health Sci Res*. 2014;4(6):889-98. DOI: 10.4103/2141-9248.144907.
26. Mannhold R, Poda GI, Ostermann C, Tetko IV. Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds. *J Pharm Sci*. 2009;98(3):861-93. DOI: 10.1002/jps.21494.
27. Paule RC, Mandel J. Consensus values, regressions, and weighting factors. *J Res Natl Inst Stand Technol*. 1989;94(3):197-203. DOI: 10.6028/jres.094.020.
28. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-10. PMID: 2868172.
29. Cochran WG. The combination of estimates from different experiments. Biometrics. 1954;10(1):101-29. DOI: 10.2307/3001666.
30. Chuang S-Y, Lin Y-K, Lin C-F, Wang P-W, Chen E-L, Fang J-Y. Elucidating the skin delivery of aglycone and glycoside flavonoids: how the structures affect cutaneous absorption. *Nutrients*. 2017;9(12):1304. DOI: 10.3390/nu9121304.
31. Giavarina D. Understanding Bland Altman analysis. *Biochem Med*. 2015;25(2):141-51. DOI: 10.11613/bm.2015.015.
32. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-63. DOI: 10.1016/j.jcm.2016.02.012.
33. Cumming H, Rücker C. Octanol-water partition coefficient measurement by a simple (1)H NMR Method. *ACS Omega*. 2017;2(9):6244-9. DOI: 10.1021/acsomega.7b01102.
34. Knutson K, Potts R, Guzek D, Golden G, McKie J, Lambert W, *et al.* Macro-and molecular physical-chemical considerations in understanding drug transport in the stratum corneum. *J Control Release*. 1985;2:67-87. DOI: 10.1016/0168-3659(85)90034-3.
35. Elias PM, Cooper ER, Korc A, Brown BE. Percutaneous transport in relation to stratum corneum structure and lipid composition. *J Invest Dermatol*. 1981;76(4):297-301. DOI: 10.1007/BF00417155.
36. Michaels AS, Chandrasekaran SK, Shaw JE. Drug Permeation through Human Skin: Theory and *in-vitro* experimental measurement. *AIChE J*. 1975;21(5):985-996. doi:10.1002/aic.690210522.
37. Flynn GL. Physicochemical Determinants of Skin Absorption. In: *Principles of Route-to-Route Extrapolation for Risk Assessment : Proceedings of the workshop held in South Carolina and North Carolina* [Gerrity TR, Henry CJ, Eds]. Elsevier; New York, NY; 1990. p. 93-127. ISBN: 9780444015822.
38. Tsakovska I, Pajeva I, Al Sharif M, Alov P, Fioravanzo E, Kovarich S, *et al.* Quantitative structure-skin permeability relationships. *Toxicology*. 2017;387:27-42. DOI: 10.1016/j.tox.2017.06.008.
39. Schrödinger Release 2020-3. QikProp, Schrödinger, LLC, New York, NY; 2020.
40. Naseem S, Zushi Y, Nabi D. Development and evaluation of two-parameter linear free energy models for the prediction of human skin permeability coefficient of neutral organic chemicals. *J Cheminformatics*. 2021;13(1):25. DOI: 10.1186/s13321-021-00503-5.
41. Johnson ME, Blankschtein D, Langer R. Evaluation of solute permeation through the stratum corneum: lateral bilayer diffusion as the primary transport mechanism. *J Pharm Sci*. 1997;86(10):1162-72. DOI: 10.1021/js960198e.
42. Mitragotri S, Anissimov YG, Bunge AL, Frasch HF, Guy RH, Hadgraft J. Mathematical models of skin permeability: An overview. *Int J Pharm*. 2011;418(1):115-29. DOI: 10.1016/j.ijpharm.2011.02.023.

PEER-REVIEWED CERTIFICATION

During the review of this manuscript, a double-blind peer-review policy has been followed. The author(s) of this manuscript received review comments from a minimum of two peer-reviewers. Author(s) submitted revised manuscript as per the comments of the assigned reviewers. On the basis of revision(s) done by the author(s) and compliance to the Reviewers' comments on the manuscript, Editor(s) has approved the revised manuscript for final publication.