DOI: 10.1002/minf.201900071

Prediction of Passive Membrane Permeability by Semi-Empirical Method Considering Viscous and Inertial Resistances and Different Rates of Conformational Change and Diffusion

Yoshifumi Fukunishi,*[a] Tadaaki Mashimo,[b, c] Takashi Kurosawa,[b, d] Yoshinori Wakabayashi,[e] Hironori K. Nakamura,[f] and Koh Takeuchi[a]

Abstract: Membrane permeability is an important property of drugs in adsorption. Many prediction methods work well for small molecules, but the prediction of middle-molecule permeability is still difficult. In the present study, we modified a classical permeability model based on Fick's law to study passive membrane permeability. The model consisted of the distribution of solute from water to membrane and the diffusion of solute in each solvent. The diffusion coefficient is the inverse of the resistance, and we examined the inertial resistance in addition to the viscous resistance, the latter of which has been widely used in permeability prediction. Also, we examined three models changing the balance between the diffusion of solute in membrane and the conformational change of solute. The inertial resistance improved the prediction results in addition to the viscous resistance. The models worked well not only for small molecules but also for middle molecules, whose structures have more conformational freedom.

Keywords: PAMPA · Fick's law · QSPR · regression model · middle molecule

1 Introduction

Permeability is one of the most important factors in a drug's adsorption and target-binding properties in cells. The understanding and predicting membrane permeability of molecules have been studied for last few decades. It is still one of the hot topics, especially under circumstances where the molecular weights of drug molecules have been increasing and larger molecules often face the lower permeability than smaller drug molecules do. There have been a number of reports on permeability.[1–26] The main permeability problems are adsorption in human intestine, extraction from kidney, penetration of the blood-brain barrier, skin permeability, and the permeability of the cell membrane to approach target proteins in cells. Caco-2 cells and MDCK cell systems are two of the model systems that mimic human intestine adsorption and extraction from kidney, respectively. Parallel artificial membrane permeability assay (PAMPA) systems have been popular in vitro assay systems for the past 20 years.[26–32] PAMPA systems have been improved to mimic in vivo permeability by trying various membrane materials, pH levels of donor and acceptor liquids and the other conditions. Certain mechanisms underlie permeability.[2,3,26] Namely, solute molecules penetrate the cell membrane by diffusion (transcellular), the solute molecules go through the tight junction (paracellular), and transporters and channel proteins work in the influx and efflux processes. Among these mechanisms, PAMPA permeability represents transcellular passive permeability.

Recent advances in molecular dynamics (MD) simulations enable the understanding and evaluation of trans-

[a] *Y. Fukunishi, K. Takeuchi*
Molecular Profiling Research Center for Drug Discovery (molprof), National Institute of Advanced Industrial Science and Technology (AIST), 2-3-26, Aomi, Koto-ku, Tokyo 135-0064, Japan
E-mail: y-fukunishi@aist.go.jp

[b] *T. Mashimo, T. Kurosawa*
Technology Research Association for Next-Generation Natural Products Chemistry, 2-3-26, Aomi, Koto-ku, Tokyo 135-0064, Japan

[c] *T. Mashimo*
IMSBIO Co., Ltd., Owl Tower, 4–21-1, Higashi-Ikebukuro, Toshima-ku, Tokyo 170-0013, Japan

[d] *T. Kurosawa*
Hitachi Solutions East Japan, 12–1 Ekimaehoncho, Kawasaki-ku, Kawasaki, Kanagawa 210-0007, Japan

[e] *Y. Wakabayashi*
BY-HEX LLP, 1–19-14, Shimizu, Suginami-ku, Tokyo 167-0033, Japan

[f] *H. K. Nakamura*
Biomodeling Research Co., Ltd., 1-704-2 Uedanishi, Tenpaku-ku, Nagoya, Aichi 468-0058, Japan

Supporting information for this article is available on the WWW under https://doi.org/10.1002/minf.201900071

cellular passive permeability.[4–15] In these calculations, MD simulations have been applied to explicit atomic models of membrane molecules with solvent water molecules. Since permeation is a very slow process, biased MD simulations have been popular in this analysis. MD simulations have shown that the distribution of the existence probability of solute and the diffusion of solute in the membrane gave the permeability constant.

On the other hand, many approaches have adopted quantitative structure-property relationship (QSPR) models for passive permeability based on Fick's law.[2,3,26] Previous works have shown the efficiency of this basic theory, and some extensions from this theory have improved prediction accuracy.[16,17] Fick's law explains that permeability is a combination of the transfer of solutes from the donor water into the membrane and the diffusion of the solutes from the donor to the acceptor sides in the membrane. Only the neutral molecule moves into the lipid layer so that the pK_a and the partition coefficient (Log$P$) or the distribution coefficient (Log$D$) determine the distribution of solute between the donor water and the membrane. The permeability P_{app} is given as follows.

$$LogP_{app} = Log\frac{D \cdot M}{h} \tag{1}$$

where P_{app}, D, M, and h represent the apparent membrane permeability, distribution coefficient, diffusion constant of the solute, and thickness of the membrane, respectively. The above MD simulations support this assumption. Namely, Log$D$ corresponds to the probability distribution of existence and M corresponds to the diffusion of solute obtained by the MD simulation, respectively.

The diffusion constant M is estimated by the Einstein-Stokes relation

$$M = \frac{k_B \cdot T}{6\pi \cdot \mu \cdot R} \tag{2}$$

where k_B, T, μ, and R are the Boltzmann constant, temperature, viscosity and radius of the solute, respectively. Since Log$D$ corresponds to the free energy of the transfer from water to membrane, this value could be approximated by the accessible surface area (ASA) method, the generalized Born (GB) method, polar surface area (PSA), and so on.[33] Thus, the following linear regression model can estimate LogP_{app}.

$$LogP_{app} = LogD - LogR + c_0 = \sum_{i=1}^{N_{descriptor}} c_1^i x_i + c_0 \tag{3}$$

where x, c, and $N_{descriptor}$ are the molecular descriptors, the fitting parameters, and the total number of descriptors, respectively, and the fitting parameters represent the characteristic properties of the membrane.

Recent advances in pharmaceutical research have increased the molecular size of drugs, and middle-molecule drugs, with molecular masses >500 Da, have become popular. In the last few decades, the major drug targets have been receptors and enzymes. Pharmaceutical companies have released several thousands of drug molecules; however, they are now facing a severe depletion in the druggable targets with conventional strategies. Research activities focusing on protein-protein interaction (PPI) inhibitors have been started instead of the projects on finding ligands to receptors and enzymes.[34,35] To inhibit PPIs, larger molecules are often preferable than the small molecules. However, the hydrophobic and often insoluble physical properties of middle molecules that are distinct from small ligands cause serious problems in their development stages.

While one of the major problems of middle molecules is the low membrane permeability, the synthetic accessibility in chemical modifications is also limited. The synthetic processes of middle molecules are more complicated and time-consuming than that of small molecules. Thus, we need the mechanistic understandings to unveil the dominant factors of membrane permeability, rather than black-box permeability predictions, to guide a rational modification of the middle molecules.

A number of reports suggested that conformational change is essential in permeability and cyclization as well as in the methylation of main-chain amide groups, and some chemical modifications have improved permeability.[18–22] Many studies have challenged this problem by using QSPR models and have successfully predicted membrane permeability, including that of middle molecules.[23] However, some studies have suggested that it is still difficult to predict the membrane permeabilities of middle molecules.

Diffusion constant M in eq. 1 consists of viscous resistance and inertial resistance. Equation 2, which is the inverse of viscous resistance, assumes the slow migration of solute in equilibrium. Permeability is non-equilibrium process in general, and the density of the solute keeps changing in the experiment. When solute molecules push solvent molecules, the inertial resistance is proportional to the cross section of the solute. In general, we have ignored inertial resistance without examining the actual experimental data. In a non-equilibrium state, the density of solute in the membrane cannot reach equilibrium. The solute can pass through the membrane before the conformational ensemble of the solute reaches equilibrium distribution (see Figure 1A).[4] In this case, the distribution of solute is different from the distribution coefficient D observed in equilibrium.

In the present study, we examined the diffusion process by using a classical QSPR model. We considered inertial resistance the same as viscous resistance. Also, we developed formulas considering the above two cases in which the conformational change occurs faster or slower than the diffusion in the membrane.[5]

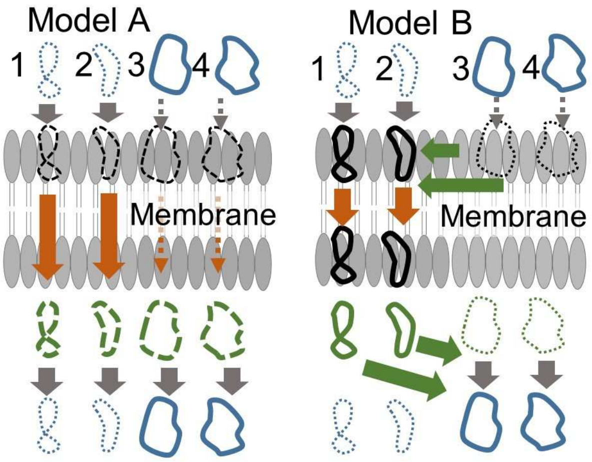

Figure 1. Permeability mechanism Models A and B. The rings represent the solute molecules and the different shapes represent the different conformers. Thick solid and thin dotted rings represent the large and small populations of molecules, respectively. The buffer solvent region is separated by the lipid bilayer in the middle (gray). Orange and green arrows represent the diffusion in the membrane and conformer change in the membrane depicted in gray. The open forms (3 and 4) are major rather than the closed form (1 and 2) in the buffer solvent. In Model A, there is no conformation change in the membrane. In Model B, the solute molecules in membrane show conformation change and the closed form becomes major.

2 Methods

We classify the membrane permeability process into two models according to the speed of permeability.[5] If the membrane permeation is faster than the conformational change of the solute in the membrane, each conformer penetrates the membrane while keeping the 3D structure of each conformer, and the dominant conformer in water could be the main contributor to the permeability (permeability model A in Figure 1A). If the membrane permeation is slower than the conformational change of the solute in the membrane, the most stable conformer in the membrane could contribute to the permeability (permeability model B in Figure 1B). In this article, we apply Fick's law to membrane permeability, since the physical meaning of Fick's law is clear and simple.

2.1 Permeability Model A

Since each solute penetrates the membrane without conformational change, the distribution of conformers in the membrane is equal to that in water solvent. The total permeability ratio ($P_{app}{}^{all}$) is the summation of the permeability of all conformers.

$$LogP_{app}^{all} = Log\left(\sum_{a=1}^{N_{conformer}} d_{wat}(a)P_{app}(a)\right) \tag{4}$$

Here, $P_{app}(a)$, $d_{wat}(a)$, and $N_{conformer}$ are the apparent permeability of the a-th conformer, the distribution of the a-th conformer in water, and the total number of conformers, respectively. The following relation is obvious.

$$P_{app}^{all} = \left\langle \frac{D(a)M(a)}{h} \right\rangle_{wat} = \left\langle exp\left(log\frac{D(a)M(a)}{h}\right)\right\rangle_{wat} \\ = \langle exp(LogP_{app}(a) \cdot log10)\rangle_{wat} \tag{5}$$

where the brackets $< >_{wat}$ stand for the average over the distribution in water. The fraction of the a-th conformer in water ($=d_{wat}(a)$) is given by

$$d_{wat}(a) = \frac{n(a)\exp(-E^{wat}(a)/(k_BT)}{\sum_{b=1}^{N_{conformer}} n(b)\exp(-E^{wat}(b)/(k_BT)} \tag{6}$$

where $E^{wat}(a)$ and $n(a)$ are the energy and the degeneracy of the a-th conformer, respectively.

When we apply Kubo's cumulant expansion to eq. 5, the first two terms of the expansion are as follows.[36]

$$LogP_{app}^{all} =< LogP_{app}(a) >_{wat} - \\ \frac{1}{2}\langle LogP_{app}(a)^2 - \langle LogP_{app}(a)\rangle^2\rangle_{wat} + \frac{1}{6}\cdots \tag{7}$$

The first term is the average LogP_{app} and the second is the deviation of LogP_{app}.

A linear regression model that is a weighted summation of the descriptor values approximates the LogP_{app} of the a-th conformer as follows,

$$LogP_{app}(a) = Log\frac{D(a)M(a)}{h} = \sum_{i=1}^{N_{descriptor}} c_1^i x_i(a) + c_0 \tag{8}$$

where c and $x(a)$ are the fitting parameters and the descriptor of the a-th conformer, respectively.

If the descriptors x_A and x_B are independent from each other, $<x_A+x_B>=<x_A>+<x_B>$ and also $\sigma(x_A+x_B)=\sigma(x_A)+\sigma(x_B)$, where σ stands for the deviation. Thus, if all the descriptors are independent of each other, eq. 7 becomes as follows.

$$LogP_{app}^{all} = c_0 + \sum_{i=1}^{N_{descriptor}} c_1^i < x_i >_{wat} + \sum_{i=1}^{N_{descriptor}} c_2^i \sigma(x_i)_{wat} + \cdots \tag{9}$$

In permeability model A, the deviations of all the descriptors, including the ASA and the radius of the solute, contribute to LogP_{app}^{all} in addition to the average values of these descriptors.

2.2 Permeability Model B

If the membrane penetration is much slower than the conformation change, the distribution of conformers reaches equilibrium in the membrane and in water. The distribution constant D is given by the partition function of the molecule as follows.

$$LogD = Log\frac{Z^{oct}}{Z^{wat}} = Log\frac{\sum_{b=1}^{N_{conformer}} n(b) \cdot exp(-E(b)^{oct}/k_BT)}{\sum_{b=1}^{N_{conformer}} n(b) \cdot exp(-E(b)^{wat}/k_BT)} \tag{10}$$

where Z^{oct}, Z^{wat}, and $n(b)$ are the partition functions in octanol and in water and the degradation number of the b-th conformer. E^{oct} and E^{wat} are the energy of the conformer in octanol and water, respectively.

The D value gives the density of molecules on the donor—-membrane interface of the membrane. The summation of diffusions of all the conformers gives the total diffusion.

$$LogP_{app}^{all} = Log\left(\frac{D}{h}\sum_{a=1}^{N_{conformer}} d_{mem}(a)M(a)\right) = LogD - Logh + Log < M(a) >_{mem} \tag{11}$$

The fraction of the a-th conformer in the membrane is given by

$$d_{mem}(a) = \frac{n(a)\exp(-E^{mem}(a)/(k_BT))}{\sum_{b=1}^{N_{conformer}} n(b)\exp(-E^{mem}(b)/(k_BT))} \tag{12}$$

As in eq. 7, here we apply Kubo's cumulant expansion to eq. 11, giving

$$LogP_{app}^{all} = LogD - Logh + < LogM(a) >_{mem} - \frac{1}{2} < LogM(a)^2 - \langle LogM(a)\rangle^2 >_{mem} + \cdots \tag{13}$$

In permeability model B, the deviations of the descriptors relating to the diffusion contribute to LogP_{app}^{all} besides the average values of the descriptors. Namely, the deviation of the radius of the solute should contribute to the prediction of LogP_{app}.

2.3 Regression and Prediction

Our physical-property prediction method was a principal component regression (PCR) with an L2 regularization term based on the molecular descriptors.[37,38] The regression model was the same as that used in our previous study. Namely, the principal component (PC) analysis projected each compound into each point in a chemical space of the PC, and a multiple linear regression was applied to the molecular coordinates in the chemical space. The principal component axes were selected to minimize the root-mean-square error (RMSE) in regression. The addition of the L2 term reduced the RMSE in the prediction.

The physical property of the i-th molecule $V(i)$ is estimated based on the molecular descriptors $\{s_i^b\}$ where b represents the b-th descriptor as follows.

$$V(i) = \sum_{j=1}^{N_{axis}} c_j \cdot p_i^j + c_0 \tag{14}$$

$$p_i^j = \sum_{b=1}^{N_{descriptor}} d_b^j \cdot (s_i^b - < s^b >) \tag{15}$$

where c_j, p_i^j, and d_b^j are the intercept parameter of the linear function, the PC vector, and the loading vector, respectively. The PC analysis of the descriptor s_i^b gives the loading vector d_b^j and the PC vector (axis) p_i^j. The values of $\{c_j\}$ are determined by multiple linear regression (MLR). N_{axis} ($N_{axis} < N_{descriptor}$) is determined so as to maximize the Q^2-value obtained by cross validation tests. The parameters are determined based on the learning set and then are used for prediction. The objective function O({c}) consists of the summation of the square error between the experimental (V^{exptl}) and calculated (V^{calc}) property values with the following regularization term:

$$O(\{c\}) = \sum_{i=1}^{N_{data}} (V_i^{exptl} - V_i^{calc}(\{c\}))^2 + \lambda \sum_{i=1}^{Naxis} (c_i)^2 \tag{16}$$

Here, the parameter set {c} is determined to minimize the O value. The weighting parameter λ is fixed to a constant ($\lambda = 0.000001$) in the present study.

We apply this regression model to P_{app}, Log$P$, Log$D$, and pK_a. Since the permeability depends on the Log$P$, Log$D$, and pK_a, the P_{app} prediction model should include the descriptors for Log$D$, Log$P$, and pK_a prediction. All the prediction models shared the same molecular descriptors. We applied 4-fold CV tests to evaluate the accuracy, and all the values in the tables were predicted in the 4-fold CVs. The CV tests showed RMSE and Q^2 values. The definitions of Q^2 and RMSE are determined as follows.

$$Q^2 = 1 - \frac{\sum_{i=1}^{N_{cmp}} (Y_{pred}i - Y_{\exp}i)^2}{\sum_{i=1}^{N_{cmp}} (Y_{\exp}i - \overline{Y_{\exp}i})^2} \quad (17)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_{cmp}} (Y_{pred}i - Y_{\exp}i)^2}{N_{cmp}}} \quad (18)$$

Here, Y_{pred} and Y_{exp} represent the predicted value in validation and the experimental Y value, respectively. In the present study, we do not compare the Q^2 values of different properties, since the variances of the experimental data (denominator in eq. 17) differed from each other among the different data sets.[39]

2.4 Regression and Prediction

Models A and B use the degeneracy and energy of the conformers in eqs. 6 and 9 so that we must sample the conformers. Although an exhaustive search of conformers is very difficult and time-consuming, many heuristic conformer generation methods have been proposed. The major approaches are random search and systematic search.[40–47] Especially, the conformer generation of ring systems is more difficult than that of nonring systems. For ring systems, some specialized methods, such as corner and edge flips, have been used.[40,41] We applied a Monte Carlo random-sampling method.[42] In conformer generation, a difficult problem is the estimation of degeneracy ($n(a)$) in eqs. 6 and 9. Most conformer generation programs focus on the reproduction of the most stable conformers precisely, but the degeneracy is unclear. Long-time MD simulations could measure the degeneracy but are not realistic in this work, since there are too many molecules.

One of the simplest approaches should be a uniform sampling of conformers without solvent bias. If the conformer search was a random sampling of dihedral angles of rotatable bonds of a molecule, the generated structural ensemble should mimic the uniform sampling (see some discussions in APPENDIX A in the supporting information). Unfortunately, this approach could not work in the sampling of large or cyclic compounds, since the ring systems could be opened by the random rotations of the dihedral angles and an atomic collision could occur. Thus, we applied a force field to close the ring systems and to avoid atomic collisions. Our conformer generation program transforms the ring systems of a molecule to nonring systems by removing one of the bonds in the ring. The following energy optimization closes the rings. The force field used was an AMBER-like one including 1–2, 1–3, 1–4, and 1–5 interactions. The 1–4 and 1–5 interaction terms consisted of only van der Waals interactions without electrostatic interactions, since the dielectric constant of water is very different from that of membrane.

A clustering analysis of the conformers generated above estimates the degeneracy number ($n(a)$ in eqs. 6 and 9). The clustering threshold is the heavy atom RMSD < 1.5 Å.

2.5 Descriptors

The molecular descriptors consisted of physical and chemical ones. The physical descriptors represent mainly the size of a molecule that is related to the diffusion. The accessible surface area (ASA) is a useful idea with which to represent the solute − solvent interaction. The chemical properties represent mainly the hydrophobicity/hydrophilicity of a molecule that is related to the distribution between water solvent and membrane. In general, charged molecules cannot penetrate a membrane whose dielectric constant is small, but neutral molecules can. The total charge of a molecule is determined by the pK_a and the MACCS key,[48] which is a set of substructures that recognizes the functional groups closely related to pK_a.[48] The atomic charges were AM1-BCC charges obtained by MOPAC7.[49–51] The hydrogen bonds are determined for the hydrogen atoms of OH and NH groups and for the acceptor atoms with lone pairs (O, N, S). The GBSA method was used to calculate the Log$D$ by eq. 10.[33] Since the descriptors include the ASA, the ASA and the Log$D$ term double-counted the solute-solvent interaction. The PCR could combine these dependent terms and thus reduce the number of independent fitting parameters. The atomic solvation parameters in the GBSA method were set to 10 cal/mol/Å2 and −5 cal/mol/Å2 for water and membrane, respectively.

Models A (eq. 9) and B (eq. 13) include both the average and deviation terms. The average $<A>$ and deviation $\sigma(A)$ of property A are determined as follows. The conformer generation program in section 2.4 generates the conformers.

$$< A >= \frac{\sum_{i=1}^{N_{conformer}} n(i)A(i)\exp(-E^{sol}(i)/(k_BT))}{\sum_{i=1}^{N_{conformer}} n(i)\exp(-E^{sol}(i)/(k_BT))} \tag{19}$$

$$\sigma(A) = \frac{\sum_{i=1}^{N_{conformer}} n(i)(A(i)- < A >)^2 exp(-E^{sol}(i)/(k_BT))}{\sum_{i=1}^{N_{conformer}} n(i)exp(-E^{sol}(i)/(k_BT))} \tag{20}$$

where sol represents the solvent (water or membrane). E^{sol} (i) is the relative energy of the i-th conformer in the solvent compared to that in vacuum.

Since membrane permeability is related to Log$D$, or to Log$P$ and pK_a values as described in eqs. 1, 3, 8, and 13, the descriptors that could approximate Log$D$, Log$P$, and pK_a values should contribute to P_{app} values. Dissociations of hydrogen atoms depend on the chemical bonds of the functional groups and the electrostatic field generated by the charge distribution of the solute. A previous work proposed a linear correlation between the proton chemical shift and the pK_a value, and the chemical shift depends on the electron density on the nucleus.[52] Thus, we adopted the descriptors used in the Log$D$ and pK_a predictions. Namely, we used the numbers of hydrogen donors and acceptors, the atomic charges of the hydrogen atoms of the NH and OH groups, and the MACCS key to represent the chemical structures of the solutes.

3. Data Preparation

The PAMPA permeability data and molecular structures were extracted from ChEMBL database version 24.[53] The ChEMBL assay and compound IDs used are summarized in Table S1. Most of the experimental conditions included pH 7.4 and the observation times were several hours, but the details of the conditions varied. In addition, a variety of membrane materials have been used.[15,26–32] A desirable assay data set consists of the P_{app} values obtained under the unique experimental conditions.[18] The careful data selection limited the number and diversity of examined compounds. While there have been differences in the P_{app} values observed through the PAMPA assay, the systems were designed to reproduce P_{app} values of the Caco-2 assay system. To evaluate the wide variety of compounds, the present data set consisted of 737 compound data obtained under the heterogeneous experimental conditions.; There were 70 pairs of the same solute with different P_{app} values; for these 70 pairs, the average difference was 0.401 and the RMSD was 0.500. Thus, we discussed the prediction accuracy of LogP_{app} >0.5. As with the P_{app}, we also prepared data on the octanol-water distribution coefficient (Log$D$, 4215 compounds),[54] and the dissociation coefficient (pK_a, 240 compounds).[55,56]

Astellas Pharmaceutical kindly provided additional experimental data, since there were fewer data on macrocyclic compounds than on acyclic compounds. The observed experimental property data were P_{app}, Log$D$, and Log$S$ data. The test compounds were 58 selected commercially available macrocyclic compounds. In addition, we asked Enamine Ltd. to evaluate the P_{app} values of two cyclic peptides using the PAMPA assay. The molecular structures and experimental values are summarized in Tables S1 and S2 in the supporting information.[57]

We prepared three-dimensional molecular structures of the examined compounds in the present study. The computational procedure was summarized as APPENDIX B in the supporting information. The initial molecular structures were 2D planar structures without hydrogen atoms in the SD file format. A file transfer program, Hgene/myPresto, was used to prepare the three-dimensional molecular structures with hydrogen atoms in electrically neutral forms including acids and amines. The atomic charges were the AM1-BCC charges provided by the MOPAC program. A molecular dynamics simulation program, cosgene/myPresto, gave the energy-optimized structures with the general AMBER force field (GAFF) assigned by the tplgeneL/myPresto topology generator.[58,59] The energy optimization gave one stable or meta-stable molecular structure for each molecule in vacuum. All the molecular structures were prepared in the Sybyl mol2 file format.

The total data set consisted of 795 data points from the ChEMBL database and the unpublished assay data. Table 1 shows the distributions of molecular weight and the total number of atoms as well as the numbers of heavy atoms, ring structures, rotatable bonds, and atoms included in the biggest ring system of each molecule.

Table 1. Statistical features of the data set for permeability prediction.

	Average	Deviation
Molecular weight (Da)	424.1	160.4
No. of all atoms	54.5	23.1
No. of heavy atoms	29.6	11.4
No. of rings	3.2	1.3

4 Results and Discussion

4.1 Prediction Models for Models A and B

Our LogP_{app} prediction model represents the dissociation of solute in water, the distribution of solute from donor water (donor) to membrane, and the diffusion of solute in the membrane towards the water (acceptor) phase. Diffusion depends on the solute molecular radius R, and many previous works adopted R as the descriptor in LogP_{app}

prediction.[2,16,23,24] The ES term (eq. 2) represents the viscous resistance, while the real solute molecule feels both the viscous resistance and the inertial resistance. The inertial resistance is proportional to the product of the cross section of the solute and the square of the velocity of the molecule. The Taylor series of the total diffusion becomes as follows.

$$\begin{aligned} LogM &= Log\left(\frac{1}{f_vR + f_iR^2}\right) \\ &= -Logf_vR(1 + f_i/f_vR) \\ &= -Logf_i - LogR + (f_i/f_v)\cdot R \\ &\quad -1/2(f_i/f_v)^2\cdot R^2 + 1/3(f_i/f_v)^3\cdot R^3 + \cdots \end{aligned} \tag{21}$$

where R, f_v, and f_i are the average radius of the solute and the respective coefficients of the diffusions related to the viscous and inertial resistance. In the actual calculation, we used the average R (denoted as $<R>$) instead of R. Thus, we described the diffusion term of LogP_{app} as a weighted summation of $<R>$, $<R>^2$, $<R>^3$, and Log$<R>$. Also, the $<R>^3$ term represented the permeant's volume, since the transfer of the molecule from the donor phase to the membrane needed a volume change of the membrane, and this change generated additional energy corresponding to pressure in the membrane. In addition to these terms, we adopted the $\sigma(R)$ term of the cumulant expansion (eqs. 7 and 13).

We discussed only a passive permeability. The Fick' law (eq. 1) suggested that the permeability process consisted of several processes in a sequential order (the diffusions in water, the partitioning between water and membrane, the diffusion in membrane) and the LogP_{app} could be given by a summation of the effects of these processes. In addition, our modifications (eqs. 7, 13 and 21) could be represented by linear regression models. Thus, we started from simple linear regression models as the QSPR equations based on Models A (eq. 22) and B (eq. 23).

$$\begin{aligned} LogP_{app}^{ModelA} &= c^D < LogD(a)_{ow}^{calc} >_{wat} \\ &+c^E\sigma(LogD(a)_{ow}^{calc})_{wat} + \sum_{i=1}^{NA} c_i^{AW}\cdot < A_i >_{wat} \\ &+\sum_{i=1}^{NA} c_i^{AM}\cdot < A_i >_{mem} + c^B\cdot < B >_{wat} + c^r\cdot < R >_{wat} \\ &+c^S\cdot < R >_{wat}^2 + c^V\cdot < R >_{wat}^3 + c^r\cdot\sigma(R)_{wat} \\ &+c^l\cdot log < R >_{wat} + c^{OH}\cdot q(OH) + c^{NH}\cdot q(NH) \\ &+c^{atom}\cdot N_{atom} + c^{rot}\cdot N_{rot} + +c^{HA}\cdot N_{HA} \\ &+c^{HD}\cdot N_{HD} + \sum c_i^{MACCS}\cdot MACCS_i + c^0 \end{aligned} \tag{22}$$

and

$$\begin{aligned} LogP_{app}^{ModelB} &= c^D\cdot Log_{10}D_{ow}^{calc} + \sum_{i=1}^{NA} c_i^{AW}\cdot < A_i >_{wat} \\ &+\sum_{i=1}^{NA} c_i^{AM}\cdot < A_i >_{mem} + c^B\cdot < B >_{mem} \\ &+c^r\cdot < R >_{mem} + c^S\cdot < R >_{mem}^2 + c^V\cdot < R >_{mem}^3 \\ &+c^r\cdot\sigma(R)_{mem} + c^l\cdot log < R >_{mem} \\ &+c^{OH}\cdot q(OH) + c^{OH}\cdot q(OH) \\ &+c^{atom}\cdot N_{atom} + c^{rot}\cdot N_{rot} + c^{HA}\cdot N_{HA} \\ &+c^{HD}\cdot N_{HD} + \sum c_i^{MACCS}\cdot MACCS_i + c^0 \end{aligned} \tag{23}$$

where A_i, B, q(OH), q(NH), N_{atom}, N_{rot}, N_{HA}, N_{HD}, and MACCS represent the ASA of the i-th atom type, the number of intramolecular hydrogen bonds, the maximum atomic charge of the H atom in the OH groups, that in the NH groups, the number of atoms, the number of rotational bonds, that of hydrogen donors, that of acceptors, and the MACCS key. The coefficient c's (c^D, c^{AW}, c^{AM}, c^B, c^R, c^S, c^V, c^r, c^l, c^{oH}, c^{NH}, c^{atom}, c^{rot}, c^{HA}, c^{HD}, c^{MACCS}, and c^0) are the fitting parameters determined by the regression. The atom type was Sybyl mol2.

4.2 Prediction Results by Models A and B: Conformer Dependence

To examine the adequacy of Models A and B, we calculated the conformer dependence of each. Before the validation of the prediction models, we examined the conformer generation. Table 2 shows the average values of $<R>_{wat}$ and $<R>_{mem}$, $\sigma(A)_{wat}$, and $\sigma(A)_{mem}$ over all of the 795 solute molecules in the dataset, when the number of conformers was set to 100. The conformer generation represented the solute-size change in water and in membrane. The results supported previous findings.[18–22,25] Namely, the solutes in membrane were folded into smaller compact structures than those in water ($<R>_{mem} < <R>_{wat}$), and the solutes in water fluctuated more than those in membrane ($\sigma(R)_{mem} < \sigma(R)_{water}$). These results showed a consensus with the previously reported phenomena.[18–22,25] Thus, we ap-

Table 2. Statistics on diffusion-related values of 795 compounds at $N_{structure} = 100$.

Term	Average	Deviation	Min	Max
$<R>_{wat}$ [a]	4.27	0.80	1.73	7.07
$<R>_{mem}$ [a]	3.85	0.75	1.73	6.44
$\sigma(R)_{wat}$ [a]	0.28	0.17	0.00	0.64
$\sigma(R)_{mem}$ [a]	0.03	0.05	0.00	0.51
$<B>_{wat}$ [b]	0.02	0.13	0.00	1.41
$<B>_{mem}$ [b]	0.02	0.16	0.00	2.00

[a]: units in Å. [b]: number of intramolecular hydrogen bonds.

plied this conformer generation method in the present study.

We applied Models A and B to the ensemble of conformers by restricting the maximum number of generated conformers up to 300. Then the 4-fold CVs of LogP_{app} predictions were used to estimate the Q^2 and RMSE values. Figure 2 and Table S3 show the conformer-number dependence of the LogP_{app} prediction. Both Models A and B worked well, and the increase in conformers improved the prediction accuracy. Some previous works showed that the conformer change increased the membrane permeability.[3,6,7,24,25] Namely, if the molecule has some conformers that show hydrophilic or hydrophobic surfaces, it can select stable conformers depending on the solvent (hydrophobic conformer in lipid and hydrophilic conformer in water).

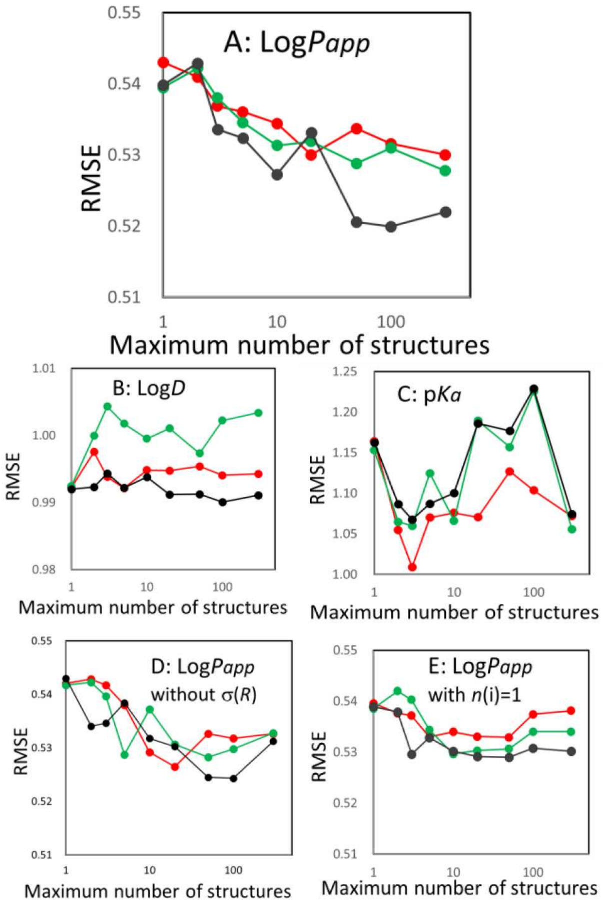

Figure 2. Conformer-number dependence of predicted properties. X axis is in Log scale. Red, green, and black circles represent the RMSE values obtained by Models A, B, and AB, respectively. (A) LogP_{app}, (B) Log$D$, (C) pK_a, (D) LogP_{app} without σ(R) terms, and (E) LogP_{app} values with $n(i) = 1$, respectively.

There are two problems that make the validation of the models difficult. One of the problems is the lack of sufficient experimental observations of the conformational dynamics of solute in permeability process. Solution NMR experiments could determine the conformations in different environments including those mimic membrane interiors. However, available NMR experimental data that directly compare aqueous and lipid bilayer environments are rather limited. Often, the limited solubilities of middle molecules prevent the observation of NMR signals in aqueous solution.

In addition, most of the permeability data are for small molecules, which are less flexible than middle molecules and have smaller number of possible conformers. There are only small number of experimentally determined permeability data on middle molecules to establish the dynamics-activity relationships. The limited number of flexible molecules with membrane permeability data also causes the problem in validating of the models based on the prediction accuracy. Recent progress of the solution NMR and molecular dynamics simulations of membrane systems should elucidate and validate the permeability mechanism in near future.

Table S4 also shows the average and maximum CPU times elapsed for one molecule. The longest CPU time was less than 62 minutes. The calculation speed should be faster than the precise MD simulation.

The results obtained by Model A were close to those by Model B. Thus, we combined the models into one by merging their descriptors. Model AB represents the united model, and the fitting parameters were determined in the same way as for Models A and B. The results obtained by Model AB were slightly better than those by Models A and B (see Figure 2 and Table S3). The RMSE values were about 0.5 and close to the deviation of the experimental data (about 0.4). Since these results showed that both Models A and B were possible permeability processes, we could not clarify the ratio of the speeds of conformer change and diffusion. The ratio should be different for each molecule in the data set.

The conformer generation represented two effects. One was the molecular-size change depending on the solvent. The radius of the molecule in water was different from that in membrane. One conformer rich of intra-molecule hydrophobic contacts could be dominant in water and the other conformer rich of intra-molecule hydrogen bonds could be dominant in water. The other effect was the structural fluctuation of the molecule. The dynamics of the molecule in water were different from those in membrane. In this case, multiple conformers existed in water and in membrane, and the population of conformers in water could be different from that in membrane. The deviation terms (σ(*R*)) in eqs. 22 and 23 correspond to this structural fluctuation of the solute molecule. To evaluate the effect of this deviation term, we removed the deviation terms from eqs. 22 and 23. Figure 2D shows the RMSE values obtained by Models A, B, and AB without the σ(*R*) terms. The σ(*R*) terms did not improve the accuracy drastically as the number of con-

Table 3. LogP_{app} prediction results obtained by various diffusion terms at $N_{structure} = 100$.

Diffusion term	Model A		Model B		Model AB	
including R	RMSE	Q^2	RMSE	Q^2	RMSE	Q^2
None	0.54	0.77	0.54	0.77	0.54	0.77
$<R>$	0.54	0.77	0.54	0.78	0.54	0.77
Log$<R>$	0.54	0.78	0.53	0.78	0.54	0.78
$\sigma(R)$	0.54	0.77	0.54	0.77	0.54	0.77
R, $\sigma(R)$	0.54	0.77	0.53	0.78	0.54	0.78
Log$<R>$, $\sigma(R)$	0.53	0.78	0.53	0.78	0.53	0.79
$<R>$, Log$<R>$	0.53	0.78	0.53	0.78	0.53	0.78
$<R>$, Log$<R>$, σ(R)	0.53	0.78	0.53	0.78	0.53	0.79
$\{<R>^n; n = 1–2\}$, Log$<R>$, $\sigma(R)$	0.53	0.78	0.53	0.78	0.53	0.79
$\{<R>^n; n = 1–3\}$, Log$<R>$, $\sigma(R)$	0.53	0.78	0.52	0.79	0.52	0.79
$\{<R>^n; n = 1–2\}$, Log$<R>$	0.53	0.78	0.54	0.78	0.53	0.78
$\{<R>^n; n = 1–3\}$, Log$<R>$	0.53	0.78	0.53	0.78	0.52	0.79

formers increased. These results suggested that the structural change of the solute molecule was important in the LogP_{app} prediction, but the dynamic structural fluctuation of the solute molecule in each condition was not so important in the LogP_{app} prediction.

We examined which terms in eqs. 22 and 23 reflect the conformer number dependence. The permeability process includes the dissociation of solutes in water and the distribution of electrically neutral solute molecules from water to membrane. These processes relate to the Log$D$, Log$P$, and pK_a values. Thus, the permeability-prediction models (eqs. 22 and 23) should predict the Log$D$, Log$P$, and pK_a values by adjusting the coefficients $\{c\}$ of eqs. 22 and 23. Figures 2B and 2 C show the conformer dependences of the predicted Log$D$ and pK_a values, respectively. The prediction models worked and the RMSE values of these predictions were similar to those previously reported. These predicted values did not show clear dependence on the number of conformers. Using the same model, only the predicted LogP_{app} values showed the conformer-number dependence among these properties. Considering eq. 1, these results suggested that the conformer-number dependence of LogP_{app} originated from the diffusion process (M in eq. 1) in the present models and that the $<R>$ terms in eqs. 22 and 23 should contribute to the conformer-number dependence of LogP_{app}.

We also examined the RMSE and Q^2 values when all the degeneracies of conformers were identical ($n(i) = 1$ in eqs. 19 and 20). Figure 2E shows the conformer dependence of the predicted LogP_{app} values with $n(i) = 1$. The results were worse than those with estimated degeneracy, and increasing the number of generated conformers did not improve the RMSE. These results suggested that the conformer generation worked properly and that the estimation of conformer populations was important in the LogP_{app} prediction.

4.3 Contribution of Diffusion Terms to LogP_{app} Prediction

We examined the diffusion process in terms of viscous (ES term, eq. 2) and inertial resistances. Table 3 shows the Q^2 and RMSE values for various combinations of diffusion terms. The diffusion was described by the Log$<R>$, $<R>$, $<R>^2$, $<R>^3$, and $\sigma(R)$ terms in eqs. 22 and 23. Since the Taylor series in eq. 21 includes higher-order terms than $<R>^3$, we examined the effect of $<R>^4$ too. $<R>^3$ was proportional to the volume of solute molecule, but it did not originally refer to volume. $<R>^2$ and $<R>^3$ correspond to the cross section of the solute molecule and the cross section of it that causes inertial resistance. The $<R>$ and Log$<R>$ terms represented viscous resistance. The $<R>^2$ and $<R>^3$ terms improved the accuracy, as did the ES term that corresponded to $<R>$ and Log$<R>$. The higher-order term ($<R>^4$) and the deviation $\sigma(R)$ did not improve the results so much.

4.4 Molecular Weight and Ring-Size Dependences

We examined how Models A and B worked in the prediction of the LogP_{app} of middle molecules. If the prediction model was adequately constructed, the predicted results should not depend on the molecular size. Since molecules with $MW > 500$ Da and those with N_{ring} (the number of ring-member atoms of the biggest ring) > 12 are so-called middle molecules and macrocyclic molecules, respectively, we examined the MW and N_{ring} dependences of the predicted data distributions. Figures 3 and 4 show the prediction results obtained by the 4-fold CV of Model B at $N_{structure} = 100$. Figure 3A and 4 A show the correlations between the predicted and experimental LogP_{app} values, and Figure 3B and 4B show the principal component analysis results as the chemical space of the data points.

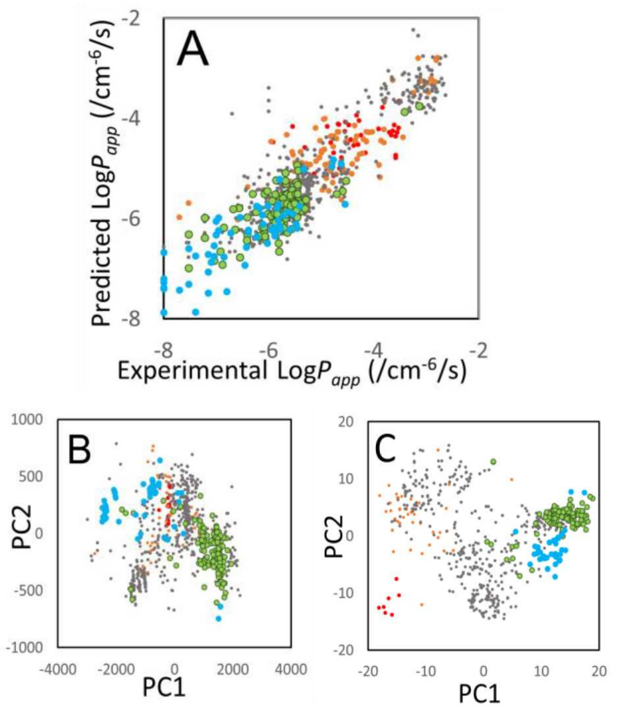

Figure 3. Predicted and experimental LogP_{app} (A), chemical spaces (B) based on the descriptors in eq. 23 and (C) based on the Mordred descriptors in term of *MW*, respectively. The model used is Model B at $N_{struct} = 100$. Red, orange, gray, green, and blue spheres represent the molecules with $0 < MW < 150$ Da, 150 Da $< MW < 300$ Da, 300 Da $< MW < 500$ Da, 500 Da $< MW < 600$ Da, and 600 Da $< MW$, respectively.

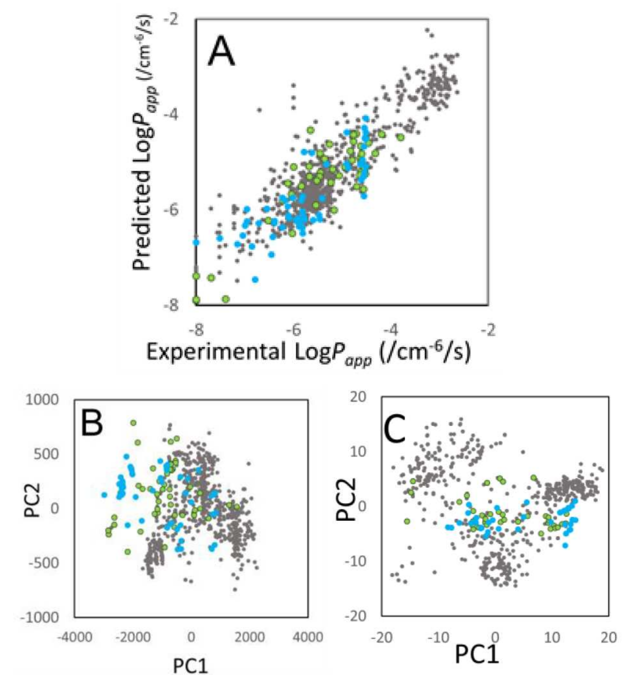

Figure 4. Predicted and experimental LogP_{app} (A) and chemical spaces (B) based on the descriptors in eq. 23 and (C) based on the Mordred descriptors, respectively. The model used is Model B at $N_{struct} = 100$. Gray, green, and blue spheres represent the molecules with N_{ring} C < 12, $12 < N_{ring} < 20$, and $20 < N_{ring}$, respectively.

Table 4. RMSE and Q^2 values in terms of molecular size at $N_{structure} = 100$.

Molecular features	No. of mols	Model A RMSE	Q^2	Model B^b RMSE	Q^2	Model AB RMSE	Q^2
$MW < 150$	37	0.47	0.36	0.45	0.43	0.46	0.40
$150 < MW < 500$	571	0.52	0.78	0.50	0.80	0.49	0.80
$500 < MW$	187	0.40	0.77	0.39	0.78	0.38	0.78
$N_{rot} < 10$	242	0.48	0.73	0.45	0.76	0.45	0.77
$10 < N_{rot} < 20$	448	0.50	0.81	0.49	0.82	0.48	0.82
$20 < N_{rot}$	99	0.46	0.84	0.43	0.86	0.44	0.85
$N_{ring} < 6$	511	0.52	0.81	0.50	0.82	0.50	0.82
$6 < N_{ring} < 11$	171	0.41	0.71	0.39	0.73	0.37	0.76
$11 < N_{ring}$	111	0.47	0.80	0.42	0.84	0.44	0.82

Since Models A and AB showed the same trends as Model B, these results were omitted.

The colors in Figure 3 represent molecular weight (*MW*). The overall trend was that the set of small molecules showed high permeability and that of large molecules showed low permeability, but the overlaps among different color points were spread across a wide area. The overlap among the smaller and bigger molecules showed wide distribution in the chemical space. The RMSE did not depend on the *MW* so much. These results suggested that the present prediction model should be robust against the *MW* difference.

The colors in Figure 4 represent the number of ring-member atoms of the biggest ring (N_{ring}). The distributions of the same color data points were not localized in the LogP_{app} prediction and chemical space, or the data points of similar size compounds spread across a wide area. The distributions of middle molecules and macrocycles overlapped that of small molecules. The RMSE did not depend on the N_{ring} so much. In addition, Table 4 shows the dependence of the RMSE on the number of rotational bonds (N_{rot}). N_{rot} represents the flexibility of a molecule. As with the *MW* and N_{ring}, the N_{rot} dependence of LogP_{app} was weak. These results suggested that the prediction model worked well in the wide molecular-size range and should be robust against various molecular sizes and shapes.

We depicted the PCA plots based on Mordred descriptors to examine the chemical space of the collected molecules in the present study.[61] Mordred consists of 1826 molecular descriptors and it has been widely used in the chemoinformatics. Figures 3C and 4 C show the results. As same as Figures 3B and 3 C, the distributions of molecules were widely spread, and the groups of molecules colored according to their size and shape distributed contiguously except very small molecules < 150 Da (colored in red). Thus, most of the data should be suitable for the present analysis..

4.5 Verifications of Models A and B

To verify Models A, B and AB, we also examined the L1 regularization (eq. 24) instead of the L2 regularization in eq. 16.

The L1 regularization could show the descriptors that show the major contributions to the results. The accuracy was almost equivalent to that obtained by the L2 regularization. Namely, Models A, B, and AB showed the RMSE and Q^2 values of 0.73 and 0.60, respectively, in the 4-fold cross validation tests (Table 5). The important descriptors suggested by the L1 regularization were the $<R>$, Log$<R>$, $<R>^2$, some ASA, mainly sub-structures including O and N atoms, and *q(OH)/q(NH)* among 324 descriptors including 166 MACCS keys. The results were summarized in Table S5 and Figure S1 in the supporting information. The results supported the results obtained by the present regression models.

Table 5. RMSE and Q^2 values obtained by eq. 24 (L1 regularization).

	Model AB	Model A	Model B
RMSE	0.59	0.59	0.59
Q^2	0.72	0.73	0.73

We examined the robustness of the present method by using hold-out tests and compared the results obtained by the Mordred-descriptor set as an alternative method. We examined five hold-out tests and compared the coefficient sets {***c***} of generated regression models, in addition to the comparisons of predicted LogP_{app} values. In each hold-out test, the molecules were sorted by the molecular features and the top 25 % bigger molecules form the hold-out set and the other smaller 75 % of molecules were used for the prediction model construction. The considered molecular features were the molecular weight (*MW*), number of atoms (N_{atom}), number of ring structures (N_{cycle}), number of rotational bonds (N_{rot}), and the number of member atoms of the maximum ring system of molecule (N_{ring}). In addition to these biased sets, we prepared a non-biased set (None) by a random selection of molecules to make a reference prediction model.

In each hold-out test, we applied the 3-fold cross validations to the teaching sets (the smaller molecules) and generated the prediction models. The prediction models estimated the LogP_{app} values of the molecules in the hold-out set (the bigger molecules). Table 6 summarizes the correlation coefficients (*R*) and the RMSE between the predicted and experimental LogP_{app} values of the hold-out set. These results show that both the present and Mordred descriptors worked well (see Figure S2 in the supporting information).

Table 6. RMSE and *R* values obtained by the hold-out tests.

Molecular feature	The present descriptors (eq. 23)			
	L1 (eq. 24)[a]		L2 (eq. 16)[a]	
	R	RMSE	R	RMSE
None[b]	0.84	0.64	0.89	0.54
MW	0.43	0.73	0.78	0.54
N_{atom}	0.81	0.78	0.90	0.58
N_{cycle}	0.45	0.62	0.59	0.53
N_{rot}	0.82	0.80	0.89	0.63
N_{ring}	0.68	0.71	0.84	0.51
Molecular feature	Mordred descriptors			
	L1		L2	
	R	RMSE	R	RMSE
None[b]	0.81	0.68	0.89	0.53
MW	0.58	0.72	0.83	0.47
N_{atom}	0.84	0.74	0.91	0.55
N_{cycle}	0.48	0.59	0.66	0.49
N_{rot}	0.84	0.76	0.90	0.61
N_{ring}	0.71	0.68	0.81	0.55

[a]: Model B was used. [b]: The reference models used in Table 7.

Table 7. Correlation coefficients (*R*) among {***c***} of the defferent prediction models in the hold-out tests.

Molecular feature	The present descriptors (eq. 23)		Mordred descriptors	
	L1 (eq. 24)	L2 (eq. 23)	L1	L2
NW	1.00	0.94	0.70	0.90
N_{atom}	1.00	0.95	0.71	0.84
N_{cycle}	1.00	0.62	0.33	0.60
N_{rot}	1.00	0.90	0.75	0.86
N_{ring}	1.00	0.88	0.81	0.63

Then, we examined the robustness of the model construction. Table 7 shows the correlation coefficients between the coefficients {***c***} of the prediction model obtained by the hold-out test and that of the reference model. The {***c***} based on the present descriptors did not depend on the difference of the teaching sets so much. On the other hand, the {***c***} based on the Mordred descriptors depended on the difference of the teaching sets strongly. The PAMPA systems represent a passive permeability only. It means that the {***c***} does not depend on the choice of the teaching sets. Thus, the present method should be realistic and useful rather than the collection of many descriptors.

5. Conclusion

We proposed a QSPR method for for evaluating the apparent membrane permeability (P_{app}) based on an analysis of the diffusion process and the partition function calculation with conformer sampling. This method generated conformers of the solute by a random structural sampling and the following structure optimization. The molecular descriptors were calculated based on the structural ensemble of the solute. Namely, the descriptors were the average and deviation values of the calculated ASA, the Log*D* and number of hydrogen bonds in water and in membrane, and the MACCS key. In addition, the descriptor set included a diffusion coefficient that is the inverse of the resistance in diffusion. To estimate the diffusion of solute, we examined the inertial resistance in the diffusion of solute in addition to the viscous resistance. We assumed two types of permeability models of solute with multiple conformers. One was the fast diffusion process (Model A), in which the solute diffused in the membrane with a fixed conformer, whose fractions were the same as those in water. The other was the slow diffusion process (Model B), in which the solute changed the conformer in the diffusion process and the fractions of conformers followed the most stable distribution in the membrane. The present QSPR models represented Models A and B based on the molecular descriptors mentioned above.

This prediction method worked in the P_{app} prediction of the middle molecules and macrocycles, the same as with that of the small molecules. The results suggested that the inertial resistance should be important in the diffusion, as is the viscous resistance known as the Einstein-Stokes equation. The ensemble of conformers improved the prediction accuracy. This study supported both Models A and B, and the permeability process could be a combination of Models A and B.

Conflict of Interest

None declared.

Supporting Information

The appendices, Tables S1–S5, Figures S1 and S2 were supplied as described in the Supporting Information.

Acknowledgements

This work was supported by grants from the National Institute of Advanced Industrial Science and Technology (AIST), by the Japan Agency for Medical Research and Development (AMED), and by the Ministry of Economy, Trade, and Industry (METI) of Japan. Astellas Pharmaceutical performed the P_{app}, Log*S*, and Log*D* experiments of macrocyclic compounds under the supervision of Dr. Tatsuya Niimi and Dr. Makoto Oku, and provided the data to us. The authors thank Prof. Haruki Nakamura and Prof. Ichio Shimada for their suggestions and helpful discussions.

References

[1] A. Finkelstein, *J. General Physiology* **1976**, *68*, 127–135.
[2] K. Bittermann, K.-U. Goss, *PLoS One* **2017**, *12*, e0190319.
[3] C. K. Wang, J. E. Swedberg, P. J. Harvey, Q. Kaas, D. J. Craik, *J. Phys. Chem. B* **2018**, *122*, 2261–2276.
[4] R. M. Venable, A. Krämer, R. W. Pastor, *Chem. Rev.* **2019**, *119*, 5954–5997.
[5] J. Witek, S. Wang, B. Schroeder, R. Lingwood, A. Dounas, H. J. Roth, M. Fouche, M. Blatter, O. Lemke, B. Keller, S. Riniker, *J. Chem. Inf. Model.* **2019**, *59*, 294–308.
[6] T. Rezai, J. E. Bock, M. V. Zhou, C. Kalyanaraman, R. S. Lokey, M. P. Jacobson, *J. Am. Chem. Soc.* **2006**, *128*, 14073–14080.
[7] T. Rezai, B. Yu, G. L. Millhauser, M. P. Jacobson, R. S. Lokey, *J. Am. Chem. Soc.* **2006**, *128*, 2510–2511.
[8] C. T. Lee, J. Comer, C. Herndon, N. Leung, A. Pavlova, R. V. Swift, C. Tung, C. N. Rowley, R. Amaro, C. Chipot, Y. Wang, J. C. Gumbart, *J. Chem. Inf. Model.* **2016**, *56*, 721–733.
[9] C. J. Dickson, V. Hornak, D. Bednarczyk, J. S. Duca, *J. Chem. Inf. Model.* **2019**, *59*, 236–244.
[10] K. Shinoda, W. Shinoda, M. Mikami, *J. Comput. Chem.* **2008**, *29*, 1912–1918.
[11] C. T. Lee, J. Comer, C. Herndon, N. Leung, A. Pavlova, R. V. Swift, C. Tung, C. N. Rowley, R. E. Amaro, C. Chipot, Y. Wang, J. C. Gumbart, *J. Chem. Inf. Model.* **2016**, *56*, 721–733.
[12] M. Badaoui, A. Kells, C. Molteni, C. J. Dickson, V. Hornak, E. Rosta, *J. Phys. Chem. B* **2018**, *122*, 11571–11578.
[13] L. W. Votapka, C. T. Lee, R. E. Amaro, *J. Phys. Chem. B* **2016**, *120*, 8606–8616.
[14] A. Kamenik, U. Lessel, J. E. Fuchs, T. Fox, K. R. Liedl, *J. Chem. Inf. Model.* **2018**, *58*, 982–992.
[15] C. H. Tse, J. Comer, Y. Wang, C. Chipot, *J. Chem. Theory Comput.* **2018**, *14*, 2895–2909, P. Matsson, J. Kihlberg, *J. Med. Chem.* **2017**, *60*, 1662–1664.
[16] R. Grohmann, T. Schindler, *J. Comput. Chem.* **2008**, *29*, 847–860.
[17] A. Cheng, D. J. Diller, S. L. Dixon, W. J. Egan, G. Lauri, K. M. Merz Jr., *J. Comput. Chem.* **2002**, *23*, 172–83.
[18] S. S. F. Leung, D. Sindhikara, M. P. Jacobson, *J. Chem. Inf. Model.* **2016**, *56*, 924–929.
[19] M. R. Sebastiano, B. C. Doak, M. Backlund, V. Poongavanam, B. Over, G. Ermondi, G. Caron, P. Matsson, J. Kihlberg, *J. Med. Chem.* **2018**, *61*, 4189–4202.
[20] G. H. Goetz, M. Shalaeva, G. Caron, G. Ermondi, L. Philippe, *Mol. Pharm.* **2017**, 14, 386–393.
[21] S. B. Rafi, B. R. Hearn, P. Vedantham, M. P. Jacobson, A. R. Renslo, *J. Med. Chem.* **2012**, *55*, 3163–3169.
[22] S. S. F. Leung, J. Mijalkovic, K. Borrelli, M. P. Jacobson, *J. Chem. Inf. Model.* **2012**, *52*, 1621–1636.
[23] P. Matsson, J. Kihlberg, *J. Med. Chem.* **2017**, *60*, 1662–1664.
[24] B. Over, P. Matsson, C. Tyrchan, P. Artursson, B. C. Doak, M. A. Foley, C. Hilgendorf, S. E. Johnston, M. D. Lee IV, R. J. Lewis, P. McCarren, G. Muncipinto, U. Norinder, M. W. Perry, J. R. Duvall, J. Kihlberg, *Nat. Chem. Biol.* **2016**, *12*, 1065–1074.
[25] M. R. Naylor, A. M. Ly, M. J. Handford, D. P. Ramos, C. R. Pye, A. Furukawa, V. G. Klein, R. P. Noland, Q. Edmondson, A. C. Turmon, W. M. Hewitt, J. Schwochert, C. E. Townsend, C. N. Kelly, M.-J. Blanco, R. S. Lokey, *J. Med. Chem.* **2018**, *61*, 11169–11182.
[26] K. Sugano, N. Takata, M. Machida, K. Saitoh, K. Terada, *Int. J. Pharm.* **2002**, *241*, 241–251.
[27] M. Kansy, F. Senner, K. Gubernator, *J. Med. Chem.* **1998**, *41*, 1007–1010.

[28] A. Avdeef, M. Strafford, E. Block, M. P. Balogh, W. Chambliss, I. Khan, *Eur. J. Pharm. Sci.* **2001**, *14*, 271–280.
[29] F. Wohnsland, B. Faller, *J. Med. Chem.* **2001**, *44*, 923–930.
[30] K. Sugano, H. Hamada, M. Machida, H. Ushio, K. Saitoh, K. Terada, *Int. J. Pharm.* **2001**, *228*, 181–188.
[31] A. Avdeef, P. Artursson, S. Neuhoff, L. Lazorova, J. Gråsjö, S. Tavelin, *Eur. J. Pharm. Sci.* **2005**, *24*, 333–349.
[32] X. Chen, A. Murawski, K. Patel, C. L. Crespi, P. V. Balimane, *Pharm. Res.* **2008**, *25*, 1511–1520.
[33] A. Daina, O. Michielin, V. Zoete, *J. Chem. Inf. Model.* **2014**, *54*, 3284–3301.
[34] L. Laraia, G. McKenzie, D. R. Spring, A. R. Venkitaraman, D. J. Huggins, *Chem. Biol.* **2015**, *22*, 689–703.
[35] H. Bruzzoni-Giovanelli, V. Alezra, N. Wilf, C.-Z. Dong, P. Tuffery, A. Rebollo, *Drug Discovery Today.* **2018**, *23*, 272–285.
[36] R. Kubo, *J. Phys. Soc. Jpn.* **1962**, *17*, 1100–1120.
[37] Y. Fukunishi, S. Yamasaki, I. Yasumatsu, K. Takeuchi, T. Kurosawa, H. Nakamura, *Mol. Inf.* **2017**, *36*, 1600013
[38] Y. Fukunishi, Y. Yamashita, T. Mashimo, H. Nakamura, *Mol. Inf.* **2018**, *37*, 1700120.
[39] D. L. J. Alexander, A. Tropsha, D. A. Winkler, *J. Chem. Inf. Model.* **2015**, *55*, 1316–1322.
[40] H. Gotō, E. Ōsawa, *J. Am. Chem. Soc.* **1989**, *111*, 8950–8951.
[41] H. Gotō, E. Ōsawa, *J. Chem. Soc. Perkin Trans.* **1993**, *2*, 187–198.
[42] A. E. Cleves, A. N. Jain, *J. Comput.-Aid. Mol. Des.* **2017**, *31*, 419–439.
[43] J. Gasteiger, C. Rudolph, J. Sadowski, *Tetrahedron Comput. Methodol.* **1990**, *3*, 537–547.
[44] A. Rusinko III, R. P. Sheridan, R. Nilakantan, K. S. Haraki, N. Bauman, R. Venkataraghavan, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 251–255.
[45] J. C. Cole, O. Korb, P. McCabe, M. G. Read, R. Taylor, *J. Chem. Inf. Model.* **2018**, *58*, 615–629.
[46] N.-O. Friedrich, C. de B. Kops, F. Flachsenberg, K. Sommer, M. Rarey, *J. Chem. Inf. Model.* **2017**, *57*, 2719–2728.
[47] Q. Wang, S. Sciabola, G. Barreiro, X. Hou, G. Bai, M. J. Shapiro, F. Koehn, A. Villalobos, M. P. Jacobson, *J. Chem. Inf. Model.* **2016**, *56*, 2194–2206.
[48] J. L. Durant, B. A. Leland, D. R. Henry, J. G. Nourse, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273–1280.
[49] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
[50] A. Jakalian, B. L. Bush, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2000**, *21*, 132–146.
[51] A. Jakalian, B. L. Bush, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2002**, *23*, 1623–1641.
[52] T. Takaoka, N. Sakashita, K. Saito, H. Ishikita, *J. Phys. Chem. Lett.* **2016**, *7*, 1925–1932.
[53] A. Gaulton, L. J. Bellis, A. P. Bentro, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, P. Overington, *Nucleic Acids Res.* **2011**, *40*, D1100–D1107.
[54] M. C. Wenlock, T. Potter, P. Barton, R. P. Austin, *J. Biomol. Screening* **2011**, *16*, 348–355.
[55] U. Zanelli, N. P. Caradonna, D. Hallifax, E. Turlizzi, J. B. Houston, *Drug Metab. Dispos.* **2012**, *40*, 104–110.
[56] F. Yoshida, J. G. Topliss, *J. Med. Chem.* **2000**, *43*, 2575–2585.
[57] P. Upadhyaya, Z. Qian, N. G. Selner, S. R. Clippinger, Z. Wu, R. Briesewitz, D. Pei, *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 7602–7606.
[58] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Compt. Chem.* **2004**, *25*, 1157–1174.
[59] Y. Fukunishi, Y. Mikami, H. Nakamura, *J. Phys. Chem. B* **2003**, *107*, 13201–13210.
[60] N. E. Tayar, A. E. Mark, P. Vallat, R. M. Brunne, B. Testa, W. F. Van Gunsteren, *J. Med. Chem.* **1993**, *36*, 3757–3764.
[61] H. Moriwaki, Y.-S. Tian, N. Kawashita, T. Takagi, *J. Cheminf.* **2018**, *10*, 4.

Received: June 19, 2019
Accepted: September 22, 2019
Published online on October 14, 2019