



Published in final edited form as:

*J Chem Inf Model.* 2019 March 25; 59(3): 1147–1162. doi:10.1021/acs.jcim.8b00648.

## Prediction of Membrane Permeation of Drug Molecules by Combining an Implicit Membrane Model with Machine Learning

Stephanie A. Brocke<sup>†</sup>, Alexandra Degen<sup>†</sup>, Alexander D. MacKerell Jr<sup>‡,§</sup>, Bercem Dutagaci<sup>\*,†</sup>, and Michael Feig<sup>\*,†</sup>

<sup>†</sup>Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA

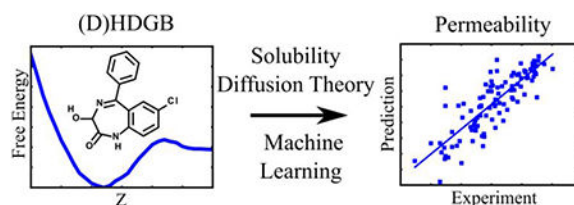
<sup>‡</sup>Department of Pharmaceutical Sciences, University of Maryland, School of Pharmacy, Baltimore, Maryland 21201, USA

<sup>§</sup>University of Maryland Computer-Aided Drug Design Center, Baltimore, Maryland 21201, USA

### Abstract

Lipid membrane permeation of drug molecules was investigated with Heterogeneous Dielectric Generalized Born (HDGB) based models using solubility-diffusion theory and machine learning. Free energy profiles were obtained for neutral molecules by the standard HDGB and Dynamic HDGB (DHDGB) to account for the membrane deformation upon insertion of drugs. We also obtained hybrid free energy profiles where the neutralization of charged molecules was taken into account upon membrane insertion. The evaluation of the predictions was done against experimental permeability coefficients from Parallel Artificial Membrane Permeability Assays (PAMPA) and effects of partial charge sets, CGenFF, AM1-BCC and OPLS, on the performance of the predictions were discussed. (D)HDGB based models improved the predictions over the two-state implicit membrane models and partial charge sets seemed to have a strong impact on the predictions. Machine learning increased the accuracy of the predictions although it could not outperform physics-based approach in terms of correlations.

### Graphical Abstract



\*Corresponding Authors: Bercem Dutagaci, 603 Wilson Road, Room BCH 218, East Lansing, MI, 48824, bercemdutagaci@gmail.com, Phone: +1 (517) 353-6743; Michael Feig, 603 Wilson Road, Room BCH 218, East Lansing, MI, 48824, mfeiglab@gmail.com, Phone: +1 (517) 432-7439, Fax: +1 (517) 353-9334.

The authors declare no competing financial interest.

SUPPORTING INFORMATION:

Tables S1-15, Figures S1-18 and Appendix S1 are provided as Supporting Material.

## INTRODUCTION

When a drug is administered to a patient, the factors of absorption, distribution, metabolism, and excretion dictate how much of the drug will reach its target inside the body, thus determining its ultimate efficacy in the body. Regardless of the route of administration (such as intravenous, oral, sublingual, topical), most drugs will need to transverse lipid bilayer membranes during transport and to enter target cells. Intravenously administered drugs would need to pass through the vascular endothelium and drugs administered orally must pass through the epithelial cells of the gut and into the bloodstream to be delivered to their target sites as well. For most small and uncharged molecules, the predominant method of membrane permeation is simply passive diffusion driven by a concentration gradient.<sup>1</sup> Thus, a critical metric for determining drug efficacy is the rate of permeation through lipid bilayer membranes.

A number of experimental methods exist for determining the membrane permeability of small molecules that are suitable in the process of drug candidate screening.<sup>2–4</sup> One such method is the Parallel Artificial Membrane Permeability Assay (PAMPA)<sup>4</sup> that utilizes an artificial membrane composed of a mixture of phospholipids embedded in a porous screen to create two compartments. In this method, a small molecule is added to one compartment and the rate of passive diffusion to the other compartment is measured. Cell-based systems for measuring the permeability coefficient such as Caco-2<sup>2</sup> and Mardin-Darby canine kidney (MDCK)<sup>3</sup> cells are also commonly used.

Experimental determination of membrane permeabilities faces limitations when large libraries of molecules need to be screened and when new molecules are considered that have not been synthesized before. As an alternative, computational methods have been proposed based on quantitative structure-permeability relationship (QSPR) models<sup>5–10</sup>, applications of solubility-diffusion theory and related kinetic models<sup>11–13</sup>, as well as the direct extraction of rates from molecular dynamics (MD) simulations<sup>14,15</sup>. The advantages of the computational methods are that permeabilities can be estimated in principle for any compound and that high-throughput protocols are feasible given sufficient resources.

QSPR models utilize a set of training compounds with known permeabilities and derive mathematical relationships between the permeability and physiochemical properties of the compounds such as the number of hydrogen bond donors/acceptors,<sup>7–9</sup> molecular and polar surface area,<sup>8,9</sup> molecular size,<sup>10</sup> and lipophilicity.<sup>9</sup> Although QSPR models are useful for filtering a large number of drug candidates, the accuracy of QSPR models is limited and such models do not perform well when a new set of drugs with different chemical properties is introduced.<sup>16</sup> QSPR models work best when the sets of test and training compounds are fairly homogeneous.<sup>17</sup> Additionally, Swift et al.<sup>18</sup> found that QSPR models have limited capacity to distinguish permeation rates for molecules that are closely related in structure.

Solubility-diffusion theory provides a fundamental framework for estimating the permeation of a given compound through a membrane bilayer.<sup>11,19</sup> This theory relates the rate of permeation of a molecule to the resistance that it experiences in the membrane according to Eq. 1:<sup>20</sup>

$$R = \frac{1}{P_m} = \int_{-d/2}^{d/2} \frac{dz}{K(z)D(z)} \quad (1)$$

where  $R$  is the resistance to permeation,  $P_m$  is the permeability coefficient (a rate, usually described in terms of cm/s),  $K(z)$  and  $D(z)$  are partition and diffusion coefficients, respectively, both of which depend on the  $z$  position of the solute along the membrane normal, and  $d$  is the membrane thickness.

In principle, solubility-diffusion theory can be applied to any molecule and any type of membrane. Its accuracy in predicting permeabilities depends primarily on the models that are used to estimate the partitioning of a given small molecule into the membrane and its interaction with the lipid bilayer as it diffuses across the membrane. Different studies have explored the application of explicit and implicit representations of membrane environment within the solubility-diffusion framework.<sup>18,21–27</sup> Generally, the idea is to determine membrane insertion free energy profiles along with estimates of intramembrane diffusion so that Eq. 1 can be applied.

The most detailed approach relies on explicit atomistic representations of the lipid bilayer. This has resulted in better agreement between computationally estimated permeation rates and experimental measurements compared to QSPR approaches.<sup>21–23, 26–28</sup> However, the accuracy of the force field, including the lipid and water parameters, are essential factors<sup>28–30</sup>. In addition, the computational costs of all-atom explicit simulations remain prohibitively high for applications to large sets of molecules.

Implicit representations of the solvent and membrane are a compromise to increase the computational efficiency of the permeation calculations. An implicit model of the membrane can rapidly estimate the partitioning of a molecule into the membrane without the need for extensive sampling<sup>31</sup>. However, implicit models typically do not allow the extraction of diffusional properties which have to be estimated separately. A comprehensive study was performed by the Jacobson group that was built upon the solubility-diffusion theory to predict the permeability of a large set of Food and Drug Administration (FDA) approved drugs and other compounds.<sup>24</sup> In that study, an implicit solvent representation based on the generalized Born (GB) method<sup>32</sup> was used to estimate the partition coefficients of the drug-like molecules between water and chloroform as an approximation of water-membrane partitioning. This work also included extensive conformational sampling to determine the lowest energy conformation of each permeant and changes in ionization state upon membrane insertion were considered. The resulting predictions correlated reasonably well with experimental permeabilities demonstrating that solubility-diffusion theory can be applied efficiently when coupled to an implicit model of the membrane layer. However, simple two-state models of the membrane environment have been critically discussed before by Marrink and Berendsen<sup>33</sup> who instead suggested a four-region model to improve accuracy. Later studies by Swift *et al.*<sup>27</sup> also commented on the possibility of oversimplification when the membrane is modeled simply by chloroform during the calculation of permeability coefficients.

In this work, we revisit the calculation of membrane permeabilities using more advanced implicit membrane models to test whether such models can lead to improved permeability estimates for drug-like molecules. More specifically, we applied the Heterogeneous Dielectric Generalized Born (HDGB)<sup>34,35</sup> model that describes the membrane via a continuously varying dielectric profile along the membrane normal along with a varying solvent accessible surface area (SASA)<sup>36</sup> term to capture electrostatic and non-polar contributions to the water-membrane partitioning free energy. In addition to the standard HDGB model, we also tested the Dynamic HDGB (DHDGB)<sup>31</sup> model that allows membrane deformations by coupling the standard HDGB model to a continuum model of elasticity.<sup>38</sup> The DHDGB framework can capture the lowering of membrane insertion free energies due to the formation of water defects for polar or charged molecules. The membrane insertion profiles from the (D)HDGB models were then applied via the solubility-diffusion framework to estimate permeability coefficients and tested on the same set of compounds that were previously studied by Leung *et al.*<sup>24</sup> With the (D)HDGB-based insertion profiles, it was possible to improve the permeability estimates over the previous work by Leung *et al.*<sup>24</sup> suggesting a benefit of the more advanced implicit membrane models. Finally, we also applied machine learning to test whether such methods can result in further improvements.

In the following, the methodology is described in more detail before results are presented and discussed.

## METHODS

### Estimation of membrane permeability coefficients

Membrane permeability coefficients ( $\log P_m$  values) were calculated based on the solubility diffusion model by calculating the integral in Eq. 1. In addition to this, we also used a simple-barrier approximation of the solubility-diffusion model according to the following expression introduced earlier<sup>39</sup> that was also used by Leung *et al.*<sup>24</sup>:

$$P_m = \frac{K_{\text{barrier}} D_{\text{barrier}}}{\delta_{\text{barrier}}} \quad (2)$$

where the membrane is treated as an effective barrier with width  $\delta_{\text{barrier}}$ .  $K_{\text{barrier}}$  is the partition coefficient of a permeant into the membrane, and  $D_{\text{barrier}}$  is its diffusion coefficient across the membrane barrier.

For the calculation of permeability coefficients via Eq. 1, partition coefficient profiles,  $K(z)$ , along the membrane normal were calculated from the free energy of membrane insertion profiles of the drug molecules using an implicit membrane model as described in more detail below.  $K(z)$  was calculated from the free energies as follows:

$$K(z) = \exp(-\Delta G(z)/RT) \quad (3)$$

where  $G(z)$  is the free energy at the specific  $z$  position. Application of Eq. 1 also requires diffusion coefficient profiles as a function of  $z$ . Viscosity profiles  $\eta(z)$  were extracted from the diffusion values reported in the studies of Carpenter *et al.*<sup>23</sup> and Bemporad *et al.*<sup>21</sup> according to the Stokes-Einstein equation:

$$D_{\text{barrier}}(Z) = \frac{k_B T}{6\pi\eta(z)r} \quad (4)$$

where  $r$  is the radius of the permeant,  $k_B$  is the Boltzmann constant, and  $T$  is the temperature. For  $z = 25 \text{ \AA}$ , the viscosity of bulk water was assumed with an experimental value of 0.0089 poise at 25°C<sup>40</sup>. We also considered a ‘high-barrier’ profile where the viscosity inside the bilayer was set to 1 poise, the value used by Leung *et al.*<sup>24</sup> All three viscosity profiles are given in Table S1. Diffusion coefficient profiles were then calculated using Eq. 4 with the radius of the permeant obtained via a spherical shape approximation from its volume for each drug molecule. The molecular volumes were calculated using two different approaches: 1) A volume based on overlapping van der Waals spheres for each atom was calculated using the CELARMM package, version c41a<sup>41</sup>; 2) the hydrodynamically effective volume was determined using HYDROPRO, version 10<sup>42</sup>.

In the simplified version of the solubility-diffusion model given in Eq. 2 we calculated  $K_{\text{barrier}}$  as follows:

$$K_{\text{barrier}} = 10^{\Delta G_{\text{barrier}}/RT - 2.3RT} \quad (5)$$

where  $G_{\text{barrier}}$  is further decomposed into the transfer free energy  $G_{\text{tr}}$  of a molecule in the conformation and charge state inside the membrane and the free energy  $G_{\text{state}}$  for converting the state preferred in water to the membrane-bound state:

$$\Delta G_{\text{barrier}} = \Delta G_{\text{tr}} + \Delta G_{\text{state}} \quad (6)$$

We generally assume that molecules are neutral inside the membrane and, therefore, at least for relatively small, rigid molecules, the main contribution to  $G_{\text{state}}$  arises from the cost of neutralizing molecules that are charged in solution.  $G_{\text{state}}$  terms were estimated using the Henderson-Hasselbalch equation based on  $\text{p}K_a$  values of ionizable sites from experimental studies.<sup>43–62</sup> We also tested the  $G_{\text{state}}$  values reported by Leung *et al.*<sup>24</sup> for the set of molecules considered here. Those values were obtained based on theoretical predictions and include an estimate of conformational reorganization free energies between water and cyclohexane as a mimic of the membrane environment.  $G_{\text{tr}}$  was extracted from membrane insertion profiles for the neutral molecules as described in more detail below. The effective width of the membrane barrier,  $\delta_{\text{barrier}}$ , was also calculated from the insertion profiles as described below.

Finally, the rate of diffusion across the barrier,  $D_{\text{barrier}}$ , was estimated based on the Stokes-Einstein equation as given in Eq. 4 without any  $z$ -dependency, where the viscosity  $\eta$  was taken as 1 poise following the study by Leung *et al.*<sup>24</sup>.

### Membrane insertion free energy profiles

Free energy insertion profiles were obtained using the HDGB and DHDGB implicit membrane models. We used the dielectric and non-polar profiles optimized for a dipalmitoyl-phosphatidylcholine (DPPC) lipid bilayer and other parameters of the GB models as described earlier<sup>34,35</sup>. Initial structures of the compounds were extracted from PubChem<sup>49</sup> in their neutral forms. The PubChem structures are minimum energy structures that resulted from extensive sampling and clustering<sup>63</sup>. The initial structures were then minimized in vacuum using CHARMM over 200 steps with a conjugate gradient algorithm. The minimized structures were positioned initially at  $z=0$ , in the center of the membrane, and subsequently translated along the  $z$  direction in 1 Å increments until  $z=30$  Å, corresponding to the bulk water phase. At each  $z$  position, the molecules were rotated around the  $x$  and  $y$  axes in 15° intervals to sample different orientations. For each position and orientation, energies were evaluated from the implicit membrane model and Boltzmann-averaged to obtain the insertion free energy as a function of  $z$  along the membrane normal. With the DHDGB model, in addition to this protocol, the thickness of the upper membrane, where the molecule was inserted, was also optimized at each  $z$  position between 0 and 25 Å to find minimum energies with locally deformed membranes.<sup>37</sup> In addition to obtaining profiles for the neutral compounds, extracting  $G_{tr}$  as described below and adding the precomputed  $G_{state}$  values, we also considered an alternative approach where we simultaneously obtained profiles for charged (at  $pH=1.4$ ) and neutral states of a given molecule. A hybrid profile was then generated that consisted of the minimum free energy from either profile after adding the cost of neutralization in aqueous solvent to the profile for the neutral form. CHARMM (version c41a1)<sup>41</sup> was used to obtain the free energy profiles. The MMTSB Tool Set<sup>64</sup> was utilized to efficiently perform the calculations on multiple molecules in parallel.

Since force field parameters were not readily available for most of the drug-like molecules studied here, the CGenFF program version 2.0<sup>65–67</sup> was used to obtain bonded and non-bonded parameters. Bonded parameters were only used for local geometry optimization. The partial charges and the choice of atomic radii that define the molecular surface were used in determining the membrane insertion profiles via the implicit membrane models. In addition to the partial charges from CGenFF, we also obtained partial charges via the AmberTools16 package<sup>68</sup> using rapid AM1-BCC quantum chemistry calculations<sup>69, 70</sup>, and OPLS2005 charges<sup>71, 72</sup> via the Ligprep module that is part of the Schrodinger Suite 2018-1<sup>73</sup>. The OPLS charges match the charge set that was used in the previous work by Leung *et al.*<sup>24</sup>. For all three charge sets, we used the atomic radii according to the Lennard-Jones  $\sigma$  parameter from CGenFF. In addition, we also tested the use of Bondi radii<sup>74</sup> with AM1-BCC charges as that combination of parameters can be obtained for new molecules without any commercial software.

We considered multiple possibilities for extracting  $G_{tr}$  and  $\delta_{barrier}$  values from the membrane insertion profiles as summarized in Fig. 1. The most straightforward approach is to estimate  $G_{tr}$  as the difference between the energy in bulk water and in the center of the membrane ( $G1$  in Fig. 1). However, this estimate neglects that there may be a significant minimum at the water-membrane interface which may increase the energy for crossing the barrier. There could also be an additional barrier upon binding to the membrane as shown in the schematic profile shown in Fig. 1. To capture these features, different estimates were considered as summarized in Fig. 1. While  $G2$  and  $G3$  are alternate estimates of  $G_{tr}$ , the quantities  $G4-9$  capture only parts of the profiles and were used as additive factors in combination with  $G2-3$ .

A simple approach for estimating  $\delta_{barrier}$  from a given profile may be to use the position of the membrane-water interface minimum as a measure of the width of the barrier ( $\delta I$ ) but other possibilities include the calculation of an effective width from different areas under the profile curve normalized by dividing by the height of the profile. Due to the symmetric nature of the membrane, the barrier should be twice as wide as the insertion profiles. Therefore, each area obtained via integration was multiplied by 2. Again, some of the  $\delta$  quantities only capture part of the profile ( $\delta6$ ,  $\delta7$ ,  $\delta8$ , and  $\delta9$ ) and those values were combined with other quantities to obtain estimates for the entire barrier (see also Table S2). The summary of all the combinations of  $G_{tr}$  and  $\delta_{barrier}$  terms tested is given in Table S2. We also used a constant  $\delta_{barrier}$  term of 18 Å since that value was used as an effective membrane thickness in previous studies<sup>24, 39</sup>. Various combinations of these terms were applied to estimate  $P_m$  using Eq. 2.

### Log $P_m$ estimates via machine learning

In addition to direct estimates of log $P_m$  values according to Eq. 1 and Eq. 2, we tested the application of machine learning algorithms that were trained on the experimental data for a subset of molecules. We discretized the DHDGB profiles along  $z$  in 2 Å increments and directly used those values along with  $G_{state}$ ,  $D_{barrier}$ , the molecular weight, and the molecular volume. Furthermore, we classified free energy profiles into different types (see Fig. 2) as a categorical feature so that a total of 20 features were input for each molecule. The classification of profiles into five classes was done based on energies in the water phase at  $z=30$  ( $E_{30}$ ), in the bilayer center at  $z=0$  ( $E_0$ ), and at the  $z$ -position where minimum energy is obtained ( $z_{min}$ ) and summarized in Table 1.

We used the Tensorflow program<sup>75</sup> and applied the deep neural network algorithm developed within this program based on the DNNLinearCombinedRegressor estimator. The Follow the Regularized Leader (FTRL) and The Adaptive Gradient Algorithm (Adagrad) optimizers were used for the linear and neural network parts of the model, respectively.<sup>76</sup> We applied the algorithm using four hidden layers with 100, 75, 50 and 25 nodes, respectively. The rectified linear unit (RELU) activation function was used for each hidden layer. The number of training epochs was set to 10,000, which was sufficient to achieve convergence in the correlations between predicted and experimental log $P_m$  values. A batch size of 4 was used, since the number of molecules considered here was relatively small. The loss of function was calculated against the experimental log $P_m$  values and the overall results



were evaluated in terms of root mean-squared errors (RMSE) and  $R^2$  correlations of linear regression curves of predicted versus experimental  $\log P_m$  values.

Machine learning was applied repeatedly for 50 molecules taken randomly as the training set. For a given training set, 10 models were generated and the model with the highest  $R^2$  value for the complete drug set was taken for that specific training set. This procedure was then repeated ten times with different randomly generated training sets to allow for a statistical evaluation of the results. The best model that gives the highest  $R^2$  with respect to experimental values is reported in more detail in the Appendix S1. The weight matrices and biases for each layer are given along with the formula for the calculation of the output.

The training sets used in machine learning were also used for multiple linear regression analysis to determine if deep learning has any advantage over much simpler multiple linear regression models. The Python statmodels module<sup>77</sup> was used for generation of the regression models. We used the same input features (discretized DHDGB profiles,  $G_{state}$ ,  $D_{barrier}$ , the molecular weight, and the molecular volume) with the machine learning calculations for the linear regression analysis to be able to directly compare machine learning and multiple linear regression approaches. The ordinary least squares method was used for estimating the regression coefficients.

### Test sets

A total of 109 molecules with available experimental PAMPA permeability coefficients ( $\log P_m$ ) were considered based on data from five different experimental studies (see structures in Figs. S1–S5).<sup>78–82</sup> These molecules are either drugs or well-characterized small chemicals and they were also studied by Leung *et al.*,<sup>24</sup> so that a direct comparison between their results and ours is possible. The three-dimensional coordinates of the drug molecules were obtained from the PubChem database<sup>49</sup>. The measured  $\log P_m$  values from these studies are reported in Table S3. In two studies, intrinsic permeability coefficients for uncharged species were reported that needed to be converted to pH dependent  $\log P_m$  values as detailed in those studies<sup>78, 80</sup>. Some of the molecules were measured in multiple studies. In these cases, we calculated average  $\log P_m$  values over multiple experimental values. The evaluation of  $\log P_m$  predictions was based on comparison against the experimental  $\log P_m$  values via linear regression to obtain  $R^2$ , slope, and intercept. In addition, RMSE values were obtained after applying an offset that minimized the RMSE values for each model. Statistical metrics for each model were calculated on the ten test sets that excluded the training sets used for machine learning to obtain statistical uncertainties and to allow for a consistent comparison with the machine learning results.

### MemDrugPerm web service

A protocol for predicting membrane permeability coefficient based on AM1-BCC charges and Bondi radii was implemented as a web server (<http://feiglab.org/memdrugperm>) to make the method available for use by the broader community. The MemDrugPerm service accepts coordinate files of small molecules as an input, automatically determines charges, calculates profiles via HDGB or DHDGB and then estimates  $\log P_m$  according to Eq. 2. Typical runs take minutes for completion.



## RESULTS

We tested the computational prediction of lipid membrane permeabilities for small drug-like molecules from free energy of insertion profiles obtained by implicit membrane models. We focused this study on a set of 109 molecules, the majority of which are FDA-approved drugs, with  $\log P_m$  data available from PAMPA experiments. The same set of compounds was also used by Leung *et al.*<sup>24</sup> so that a direct comparison with this earlier work can be made. As described in the Methods section, we estimated permeabilities based on solubility-diffusion theory, either directly via Eq. 1 or using a simplified version where the membrane is treated as a simple barrier and  $P_m$  is obtained via Eq. 2. In addition, we also tested a machine learning approach using features of the insertion profiles and selected molecular properties as input but without making any assumptions about the specific functional form for estimating  $P_m$ .

### Membrane insertion profiles

As detailed in the Methods section, membrane insertion profiles were obtained for all of the compounds with the HDGB and DHDGB implicit membrane models using CHARMM CGenFF, Amber AM1-BCC, and OPLS charge sets and atomic radii either derived from CHARMM Lennard-Jones or Bondi radii.

The membrane profiles for all of the compounds are shown in Figs. S6–S15. They can be classified into five major types (see Fig. 2). For the majority of compounds, membrane insertion is unfavorable, resulting in a significant barrier in the center of the membrane (classes 1, 2, and 4). A subset of compounds was more favorable inside the membrane than in water (classes 3 and 5). For most compounds except for the molecules in class 4 there was a small barrier upon membrane insertion and except for the small number of compounds in classes 4 and 5, the insertion profiles exhibited a free energy minimum in the membrane head-group region.

As a typical example, the insertion profiles for caffeine (class 1 or 2 depending on the model and charge set) are shown in Fig. 3. As expected, the estimated insertion free energies are significantly reduced with the DHDGB model over HDGB since DHDGB allows the deformation of the membrane to facilitate the diffusion of polar compounds. This is generally the case for the compounds with profiles in classes 1, 2, and 4 (see Figs. S6–S15). The example profiles for caffeine highlight furthermore that different charge sets can also result in significantly different insertion profiles. The choice of atomic radii only has also an effect as the insertion profiles with Bondi radii and AM1-BCC charges generally give larger barriers for the membrane insertion compared to the CHARMM Lennard-Jones radii and AM1-BCC charges. The significant quantitative variations either due to the choice of the implicit solvent model or the charge set have a direct impact on the accuracy of the  $\log P_m$  predictions as discussed below. Moreover, in some cases, even the qualitative profile classification varied as a function of the methodology (see Table S4).

### Prediction of log $P_m$ values by integration over the membrane insertion profiles

Log $P_m$  values were obtained by direct integration of the free energy and diffusion coefficient profiles following Eq. 1 based on the free energy profiles of neutral molecules obtained by the HDGB and DHDGB models. We also tested hybrid free energy profiles for charged compounds by taking into account the free energy profiles of neutral and charged molecules offset by  $G_{state}$  based on the  $pK_a$  as described in the Methods section. Diffusion coefficient profiles were obtained according to the three viscosity profiles given in Table S1 and Eq. 4 (see Methods). In addition, we also tested setting  $D(z)$  simply to the estimated value in water for any value of  $z$ . Table 2 shows the results for the neutral free energy profile and Tables 3 and 4 show the results for the hybrid free energy profiles generated using  $G_{state}$  terms obtained using experimental  $pK_a$  values and theoretical predictions from Leung *et al.*, respectively. Individual log $P_m$  values using DHDGB profiles and high-barrier diffusion profiles are listed in Tables S5–7. Generally, good correlation is found between the estimated and experimental log $P_m$  values with  $R^2$  values ranging between 0.37 and 0.60 for linear regression fits. The slopes of the regression functions were generally larger than 1 and root mean-squared errors after adding a constant offset were at best 1.3, but typically larger. Across all parameter variations, DHDGB resulted in significantly better performance than HDGB with higher correlation coefficients, slopes closer to 1, and smaller RMSE values. There were also clear trends with different charge sets. Overall, OPLS charges resulted in the highest correlations and smallest RMSE values, followed by estimates based on AM1-BCC and CGenFF charges. There was also a significant effect of the radii used with the AM1-BCC charges. With Bondi radii, the performance was worse than with the CGenFF-derived radii, especially when DHDGB was used to obtain profiles.

The use of hybrid profiles led to modest improvements in correlation, especially with DHDGB, but slopes and RMSE values were increased as well (see Tables 3 and 4). Using  $G_{state}$  values from Leung *et al.* that take into account conformational changes resulted in slightly better predictions than the use of  $G_{state}$  values estimated only based on  $pK_a$  values for OPLS and AM1-BCC charges but not CGenFF charges.

The choice of the diffusion profile had only a small effect on the results. The high-barrier profile resulted in same  $R^2$  values but also slightly increased slopes and RMSE values, but the overall differences were small indicating that the exact shape of  $D(z)$  is not critical. Even assuming a constant diffusion value from water throughout the membrane did not make much of a difference.

### Prediction of log $P_m$ values by simple barrier approximation

Log $P_m$  values were also determined according to Eq. 2 based on single values of  $K_{barrier}$ ,  $\delta_{barrier}$ , and  $D_{barrier}$ . The values for  $K_{barrier}$  and  $\delta_{barrier}$  were obtained based on the profiles (see Fig. 1) either with or without including  $G_{state}$  values according to Eq. 6. The diffusion across the membrane,  $D_{barrier}$ , was estimated either based on the molecular volume or the effective hydrodynamic volume. The individual  $D_{barrier}$  and  $G_{state}$  estimates for the molecules are given in Table S8. For the DHDGB profiles, different estimates for  $G_{tr}$  ( $G_{tr}$  1–9, see Fig. 1)) are given in Tables S9–S11 and the different estimates for  $\delta_{barrier}$  are given in Tables S12–S14.

Many combinations of  $G_{tr}$  and  $\delta_{barrier}$  were explored (see Table S2) in the context of Eq. 2 to predict  $\log P_m$ . Selected combinations leading to the highest correlations and lowest RMSE values with DHDGB are listed in Table S15. The best results were found when

$G_{barrier}$  was calculated as  $G1$ , i.e. the difference in energy between the center of the membrane and the water phase, and  $\delta_{barrier}$  as  $\delta2$ , which corresponds to the area under the entire profile divided by the energy in the center of the membrane (see Fig. 1). We note that using a constant  $\delta_{barrier}$  term as in Leung *et al.* yields almost equal  $R^2$  values as with  $\delta2$ , but the RMSE values are lower using  $\delta2$ . The inclusion of  $G_{state}$  estimates improved the correlations but led to increased RMSE values as in the direct integration according to Eq. 1. Slightly higher  $R^2$  values were obtained when  $D_{barrier}$  was calculated using the volume obtained from CHARMM instead of the hydrodynamic volume from HYDROPRO.

The  $\log P_m$  values obtained with  $G_{barrier}=G1$   $\delta_{barrier}=\delta2$  are given in Tables S5–7 and were further analyzed. Table 5 summarizes the agreement between the predicted  $\log P_m$  values and the PAMPA data. The results are similar although overall somewhat worse than the estimates obtained by direct integration via Eq. 1. Again, the profiles from DHDGB give mostly better results than with HDGB and the best and worst agreement is found with the OPLS and CGenFF charges, respectively.

We also compared our results with the previous work by Leung *et al.*<sup>24</sup> that was based on a two-state implicit model rather than z-dependent insertion profiles as in the present study. Comparing the results obtained with the OPLS charges, we obtain higher correlation coefficients, smaller slopes and lower RMSE values with HDGB and DHDGB. Since the model based on Eq. 2 was the same between our work and the work by Leung *et al.* this demonstrates that a more detailed implicit membrane formalism directly translates into better estimates of permeabilities.

### Prediction of $\log P_m$ values by machine learning

We applied machine learning algorithms for  $\log P_m$  predictions based on discretized profiles as input features as described in the Methods section, with the best resulting model then used to predict  $\log P_m$  values. Since we used part of the molecules as training data, we calculated  $R^2$ , slope, and RMSE values only for the remaining molecules that were not included in the training set. This analysis was repeated for a number of different sets of molecules selected as training and test sets. The averages over test sets are reported in Fig. 4. Since we did not test machine learning with the AM1-BCC/Bondi combination, the corresponding results with DHDGB-based predictions are not shown here. In addition to the machine learning model, we applied a multiple linear regression model for the same training sets used in machine learning to be able to compare multiple linear fits with the deep learning models.

In terms of correlation coefficients, the model based on machine learning outperforms the estimates based on Eq. 2 and based on Eq. 1 without the hybrid profiles for the CGenFF and AM1-BCC charge sets but matches the results based on Eq. 1 with the hybrid profiles for those charge models. With OPLS charges, the correlations based on the machine learning model were not significantly improved over the estimates based on Eq. 2 and worse than the estimates based on Eq. 1. The machine learning model had significantly lower RMSE

values. However, lower RMSE values can also be obtained by simply applying a linear transformation of the  $\log P_m$  predictions which would not affect the correlation coefficients. For example, for the estimates with DHDGB for OPLS charges based on Eq. 1, a simple linear transformation that scales the values so that the slope in the regression analysis with respect to the PAMPA data becomes 1 results in RMSE values of 0.76, similar to the values obtained from the machine learning. Multiple linear regression models give  $R^2$  correlations smaller than machine learning models suggesting that machine learning with deep components has advantages over the linear model.

The advantage of machine learning protocols is that no *a priori* model is required. It is remarkable that the performance achieved via machine learning is similar to a direct implementation of Eq. 1 but, perhaps somewhat disappointingly, we did not find that the machine learning model could significantly outperform the best direct estimates based on Eq. 1.

## DISCUSSION AND CONCLUSIONS

In this study, we applied HDGB based implicit membrane models and their combination with solubility-diffusion theory and machine learning to predict membrane permeabilities of drug molecules. The results were evaluated against the experimental PAMPA  $\log P_m$  values in terms of  $R^2$  correlations and RMSE values. The general finding is that the HDGB-based models result in improved predictions over previous efforts based on a two-state water/chloroform model that was used in the work of Leung *et al.* We also find that including membrane deformations via the DHDGB model leads to consistent further improvements as it is expected to better represent the insertion profiles of polar molecules. The availability of insertion profiles instead of single-value water-membrane partition coefficients allowed us furthermore to directly apply the integral form of the solubility-diffusion theory instead of the simpler barrier approximation. This also resulted in further improvements, especially when the hybrid profiles were applied that take into account changes in protonation states when charge molecules are inserted into the membrane.

The use of DHDGB resulted in overall improvements not just because of the more accurate physics but also because it allowed a compensation of overestimated charges. An example is given in Fig. 5 for sulfasalazine where the HDGB insertion free energies were much higher with CGenFF and AM1-BCC than for OPLS charges. DHDGB profiles in all of the cases were much more similar to each other and resulted in estimated  $\log P_m$  values much closer to the experimental data. This explains why the use of DHDGB resulted in overall larger improvements for estimates based on CGenFF and AM1-BCC charge sets where the initial correlations with HDGB were significantly lower than with the OPLS charges.

Even with DHDGB, there are differences with different charge sets. Overall, OPLS gives the highest  $R^2$  correlations, although all three charge sets give similar RMSE values with respect to the experimental values. This suggests that the choice of partial charges has an impact on the calculations and further improvements in the prediction of permeabilities may depend primarily on better charges. To understand the differences in the partial charges better, we illustrate in Fig. 6 three examples of drugs that have large differences in partial charges

between CGenFF and OPLS and as a result differ significantly in their free energy of membrane insertion profiles. Polar regions especially around nitrogen atoms gave the largest differences in partial charges between the two charge sets. There is a general tendency for CGenFF to overestimate the hydrophilicity for these three drugs and as a result, OPLS gave better predictions. Additional analysis of the CGenFF parameters and the charge penalties that the CGenFF program outputs shows a correlation between the penalties and the differences of the predictions and experimental values (see Fig. S16–18). The highest penalties were given for the polar regions, reflecting the overestimation of polarity that caused the larger negative differences for the  $\log P_m$  values. While knowing that charges may be inaccurate is helpful, it remains a significant challenge to actually improve the accurate estimation of partial charges, especially in a high-throughput context where large numbers of molecules need to be screened.

The HDGB and DHDGB models were optimized based on a DPPC membrane while the experimental  $\log P_m$  values are based on the PAMPA experiments with different lipid compositions, namely brain membrane<sup>81</sup>, gastrointestinal track lipids<sup>78–80</sup> or lecithin from egg yolk<sup>82</sup>. All of these membranes have fatty acid compositions with mostly unsaturated lipids with 18 or fewer carbons<sup>83–85</sup>. The direct comparison with PAMPA experiments is fundamentally problematic because PAMPA does not directly measure permeation across pure lipid bilayers. Instead, in PAMPA experiments, lipids are dissolved in organic solvent and may form a heterogeneous mixture of bilayers, liposomes, and micelles. Nevertheless, under the premise that PAMPA experiments do approximate pure bilayers, we explored whether an adjustment of the membrane thickness in the implicit membrane model could lead to better agreement with the PAMPA-based  $\log P_m$  values. Membranes formed by the lipids used in the PAMPA experiments are expected to result in slightly thinner than pure DPPC bilayers<sup>86</sup>. Therefore, we tested implicit membranes with different thicknesses by simply assuming that the shape of the insertion profiles would remain the same and that the profiles would be shifted along  $z$  in either direction with the value at  $z=0$  either repeated (for a wider membrane) or deleted (for a thinner membrane). Based on the modified profiles, we recalculated  $\log P_m$  values using Eq. 1 (see Table 6). We found that increasing the width leads to increasing RMSE values, whereas a decreased width reduced RMSE values. The better agreement with the thinner profiles is consistent with the, on average, shorter lipids used in the PAMPA experiments compared to DPPC bilayers. The analysis demonstrates more generally that the computational  $\log P_m$  estimates are sensitive to the membrane thickness and that it may be possible in the future to provide different permeability estimates for different membrane types with appropriately reparametrized (D)HDGB models.

Different lipid contents and concentrations of lipids are known to cause variations in experimental  $\log P_m$  values from PAMPA<sup>87</sup>. Table S3 shows that the experimental values for the same drugs from different experiments vary and average experimental  $\log P_m$  values have uncertainties between 0.1–0.8. One way to correct for the systematic differences is to correct  $\log P_m$  values of one data set with respect to another data set, simply by applying a linear transformation based on the equation of  $y = ax + b$ . We did so using the largest Fujikawa *et al.* set as the reference and rescaling the data sets from Avdeef *et al.* ( $a = 0.15$ ;  $b = -4.06$ ), Balimane *et al.* ( $a = 1.15$ ;  $b = 0.79$ ) and Di *et al.* ( $a = 1.22$ ;  $b = 1.94$ ) with respect to the Fujikawa set based on a regression analysis for shared compounds. The Bermejo *et al.* set

has only one common drug with Fujikawa, therefore we used the original  $\log P_m$  values of that set without scaling. Using the scaled  $\log P_m$  values as a reference, we reevaluated the predicted  $\log P_m$  values obtained by integration of the profiles from DHDGB model and using OPLS charge set (see Table 7). We find an increased correlation for both neutral and hybrid profiles. This suggests that computational  $\log P_m$  predictions may be improved further by fully understanding how PAMPA-based  $\log P_m$  predictions depend on lipid contents and concentration and, ultimately, how such experiments relate to passive diffusion across biological bilayers.

Alternatively, QSPR models could be applied to predict  $\log P_m$  permeabilities. The general idea of QSPR modeling is to fit simple mathematical models based on physicochemical descriptors to reproduce known  $\log P_m$  values for a given training set and then apply such models to new compounds. The assumption underlying QSPR modeling is that the combination of molecular descriptors can implicitly capture the factors giving rise to  $\log P_m$  values but without invoking a particular physical theory that links molecular features to permeability. In contrast, the approach taken here directly estimates the energetics and kinetics of molecules diffusing through a lipid bilayer so that solubility-diffusion theory can be applied to calculate  $\log P_m$  values.

For different sets of compounds, QSPR models were reported giving  $R^2$  correlations as high as 0.5 to 0.8<sup>7, 10, 88</sup>. However, in these studies, the training and test sets were fairly similar. For the diverse set of compounds studied here, Leung *et al.*, evaluated the correlation of experimental  $\log P_m$  values with individual physicochemical descriptors such as molecular weight, volume, or polar surface area (PSA)<sup>24</sup>. In addition, the QSPR-type regression-based models QPlog $P_{o/w}$  (octanol/water partition coefficients), QPPCaco (predicted Caco-2) and QPPMDCK (predicted MDCK) were tested. The finding was that individual physicochemical descriptors generally failed to be correlated significantly with the experimental  $\log P_m$  values, except for PSA ( $R^2=0.29$ ), whereas the trained QSPR models showed overall correlations ranging from 0.19 to 0.23 for the complete data set considered here based on the values reported in the study by Leung *et al.*<sup>24</sup>. As discussed before<sup>24</sup>, this suggests that the physics-based models applied here offer greater transferability in predicting  $\log P_m$  values for a wider range of compounds.

The protocol presented here starts with a single conformation (obtained from PubChem) and applies a deterministic protocol for obtaining membrane insertion profiles from which  $\log P_m$  values are then estimated. Therefore, there are no statistical uncertainties for the reported insertion profiles. Because no extensive sampling is involved, the protocol presented here is computationally efficient and can be applied to a large number of molecules. However, we neglect the conformational dynamics of the studied molecules and, in particular, potential differences in conformational sampling between the aqueous and membrane phases. Full-scale molecular dynamics simulations of small molecule diffusion across lipid bilayers are possible and could provide a full account of conformational dynamics<sup>21</sup>. However, the costs of such an approach would be prohibitive in the context of a practical protocol for the rapid estimation of  $\log P_m$  values. At the same time, the overall contribution of conformational dynamics may actually be quite small according to the study by Leung *et al.*<sup>24</sup>



In summary, we present an improved method for the estimation of membrane permeabilities of drug-like molecules. Overall, the highest correlation with the experimental  $\log P_m$  values was obtained by directly evaluating the integral from solubility-diffusion theory in Eq. 1 based on the hybrid profiles obtained with the DHDGB model and the OPLS partial charge set. The method is physics-based as it builds on solubility-diffusion theory using implicit membrane-based insertion free energy profiles. The method is generally applicable to a wide variety of molecules and sufficiently fast for handling large sets of molecules. A web server implementation is available for public use. A critical feature remains the estimation of partial charges where future efforts should be focused.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We thank Dr. Afra Panahi for her help with using the DHDGB model, Dr. Matthew P. Jacobson for help in validating the OPLS partial charges. SB thanks Michigan State University for support in form of a professorial assistantship and the Department of Biochemistry & Molecular Biology at Michigan State University for an Undergraduate Research Fellowship. Schrodinger Suite 2018-1 was used with the license of Institute of Cyber-Enabled Research / High Performance Computing Cluster (ICER/HPCC) at the Michigan State University.

### Funding Sources

NIH R35 GM126948, NSF MCB 1330560, NIH GM070855 (to ADM)

## ABBREVIATIONS

<b>Adagrad</b>	Adaptive Gradient Algorithm
<b>CHARMM</b>	Chemistry at Harvard Molecular Mechanics
<b>DHDGB</b>	dynamic heterogeneous dielectric generalized Born
<b>DPPC</b>	Dipalmitoyl-phosphatidylcholine
<b>FDA</b>	Food and Drug Administration
<b>FTRL</b>	Follow the regularized leader
<b>GB</b>	generalized Born
<b>HDGB</b>	heterogeneous dielectric generalized Born
<b>MDCK</b>	Mardin-Darby canine kidney
<b>MMTSB</b>	Multiscale Modeling Tools in Structural Biology
<b>PAMPA</b>	Parallel Artificial Membrane Permeability Assay
<b>QSPR</b>	quantitative structure-permeability relationship
<b>RELU</b>	Rectified Linear Unit; RMSE Root Mean Squared Error



## SASA

## Solvent-Accessible Surface Area

## REFERENCES

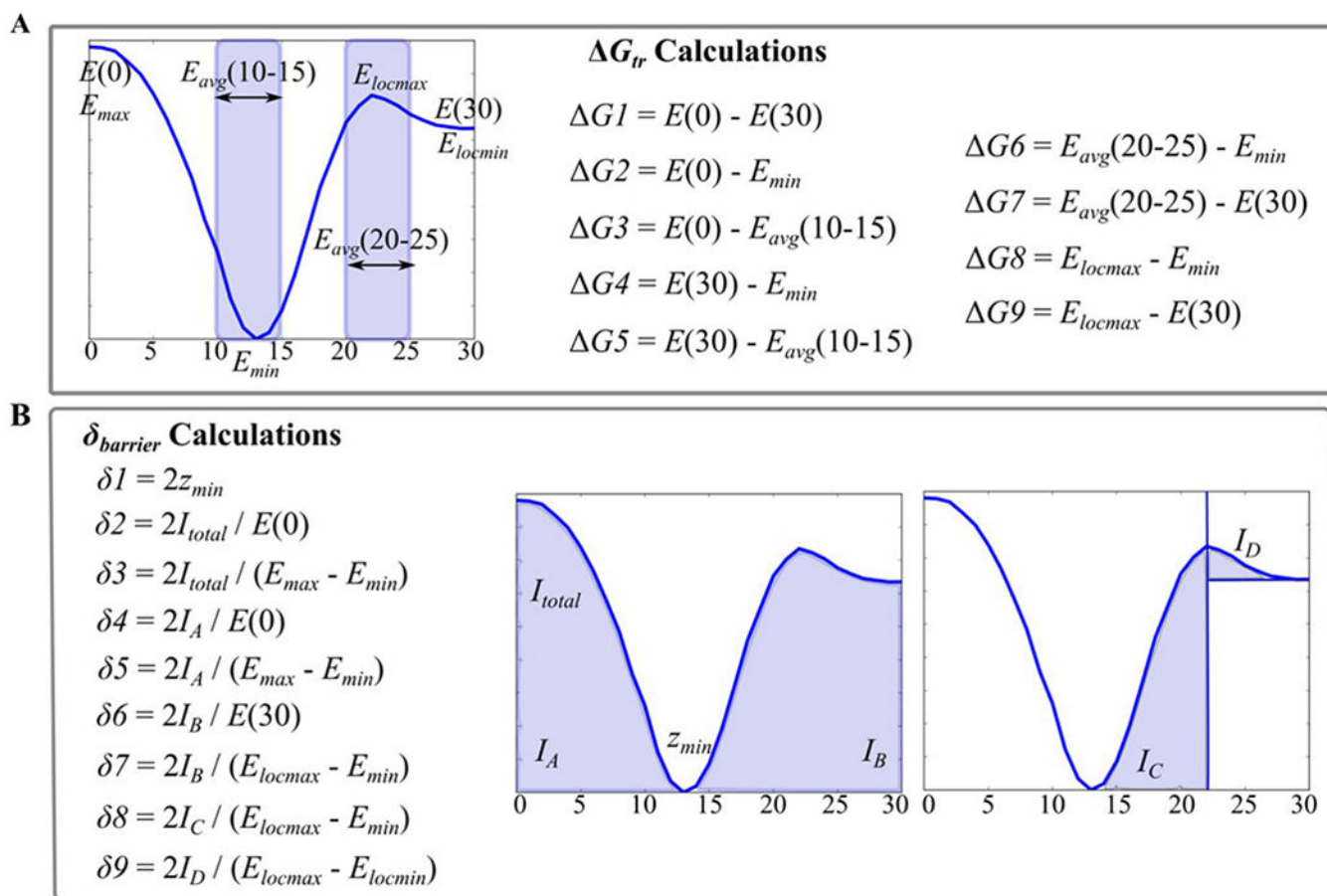
1. Seddon AM; Casey D; Law RV; Gee A; Templer RH; Ces O, Drug Interactions with Lipid Membranes. *Chem. Soc. Rev* 2009, 38, 2509–2519. [PubMed: 19690732]
2. Artursson P; Karlsson J, Correlation between Oral-Drug Absorption in Humans and Apparent Drug Permeability Coefficients in Human Intestinal Epithelial (Caco-2) Cells. *Biochem. Biophys. Res. Comm* 1991, 175, 880–885. [PubMed: 1673839]
3. Irvine JD; Takahashi L; Lockhart K; Cheong J; Tolan JW; Selick HE; Grove JR, MDCK (Madin-Darby canine kidney) Cells: A Tool for Membrane Permeability Screening. *J. Pharm. Sci* 1999, 88, 28–33. [PubMed: 9874698]
4. Kansy M; Senner F; Gubernator K, Physicochemical High Throughput Screening: Parallel Artificial Membrane Permeation Assay in the Description of Passive Absorption Processes. *J. Med. Chem* 1998, 41, 1007–1010. [PubMed: 9544199]
5. Egan WJ; Merz KM; Baldwin JJ, Prediction of Drug Absorption Using Multivariate Statistics. *J. Med. Chem* 2000, 43, 3867–3877. [PubMed: 11052792]
6. Ekins S; Waller CL; Swaan PW; Cruciani G; Wrighton SA; Wikel JH, Progress in Predicting Human ADME Parameters in Silico. *Journal of Pharmacological and Toxicological Methods* 2000, 44, 251–272. [PubMed: 11274894]
7. Oja M; Maran U, Quantitative Structure-Permeability Relationships at Various pH Values for Acidic and Basic Drugs and Drug-Like Compounds. *Sar and Qsar in Environmental Research* 2015, 26, 701–719. [PubMed: 26383235]
8. Refsgaard HHF; Jensen BF; Brockhoff PB; Padkjaer SB; Guldbrandt M; Christensen MS, In Silico Prediction of Membrane Permeability from Calculated Molecular Parameters. *J. Med. Chem* 2005, 48, 805–811. [PubMed: 15689164]
9. Winiwarter S; Bonham NM; Ax F; Hallberg A; Lennernas H; Karlen A, Correlation of Human Jejunal Permeability (in Vivo) of Drugs with Experimentally and Theoretically Derived Parameters. A Multivariate Data Analysis Approach. *J. Med. Chem* 1998, 41, 4939–4949. [PubMed: 9836611]
10. vandeWaterbeemd H; Camenisch G; Folkers G; Raevsky OA, Estimation of Caco-2 Cell Permeability Using Calculated Molecular Descriptors. *Quantitative Structure-Activity Relationships* 1996, 15, 480–490.
11. Diamond JM; Katz Y, Interpretation of Nonelectrolyte Partition-Coefficients between Dimyristoyl Lecithin and Water. *Journal of Membrane Biology* 1974, 17, 121–154. [PubMed: 4407798]
12. Walter A; Gutknecht J, Permeability of Small Nonelectrolytes through Lipid Bilayer-Membranes. *Journal of Membrane Biology* 1986, 90, 207–217. [PubMed: 3735402]
13. Nagle JF; Mathai JC; Zeidel ML; Tristram-Nagle S, Theory of Passive Permeability through Lipid Bilayers. *J. Gen. Physiol* 2008, 131, 77–85. [PubMed: 18166627]
14. Dickson CJ; Hornak V; Pearlstein RA; Duca JS, Structure-Kinetic Relationships of Passive Membrane Permeation from Multiscale Modeling. *J. Am. Chem. Soc* 2017, 139, 442–452. [PubMed: 27951634]
15. Oruc T; Kucuk SE; Sezer D, Lipid Bilayer Permeation of Aliphatic Amine and Carboxylic Acid Drugs: Rates of Insertion, Translocation and Dissociation from MD Simulations. *Phys. Chem. Chem. Phys* 2016, 18, 24511–24525. [PubMed: 27539552]
16. Stouch TR; Kenyon JR; Johnson SR; Chen XQ; Doweyko A; Li Y, In Silico ADME/Tox: Why Models Fail. *J. Comput. Aid. Mol. Des* 2003, 17, 83–92.
17. Dudek AZ; Arodz T; Galvez J, Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review. *Combinatorial Chemistry & High Throughput Screening* 2006, 9, 213–228. [PubMed: 16533155]
18. Swift RV; Amaro RE, Modeling the Pharmacodynamics of Passive Membrane Permeability. *J. Comput. Aid. Mol. Des* 2011, 25, 1007–1017.

19. Paula S; Volkov AG; VanHoek AN; Haines TH; Deamer DW, Permeation of Protons, Potassium Ions, and Small Polar Molecules through Phospholipid Bilayers as a Function of Membrane Thickness. *Biophys. J* 1996, 70, 339–348. [PubMed: 8770210]
20. Berendsen HJC; Marrink SJ, Molecular-Dynamics of Water Transport through Membranes - Water from Solvent to Solute. *Pure and Applied Chemistry* 1993, 65, 2513–2520.
21. Bemporad D; Essex JW; Luttmann C, Permeation of small molecules through a lipid bilayer: A computer simulation study. *J. Phys. Chem. B* 2004, 108, 4875–4884.
22. Bennion BJ; Be NA; McNerney MW; Lao V; Carlson EM; Valdez CA; Malfatti MA; Enright HA; Nguyen TH; Lightstone FC; Carpenter TS, Predicting a Drug's Membrane Permeability: A Computational Model Validated With in Vitro Permeability Assay Data. *J. Phys. Chem. B* 2017, 121, 5228–5237. [PubMed: 28453293]
23. Carpenter TS; Kirshner DA; Lau EY; Wong SE; Nilmeier JP; Lightstone FC, A Method to Predict Blood-Brain Barrier Permeability of Drug-Like Compounds Using Molecular Dynamics Simulations. *Biophys. J* 2014, 107, 630–641. [PubMed: 25099802]
24. Leung SSF; Mijalkovic J; Borrelli K; Jacobson MP, Testing Physical Models of Passive Membrane Permeation. *J. Chem. Inf. Model* 2012, 52, 1621–1636. [PubMed: 22621168]
25. Leung SSF; Sindhikara D; Jacobson MP, Simple Predictive Models of Passive Membrane Permeability Incorporating Size-Dependent Membrane-Water Partition. *J. Chem. Inf. Model* 2016, 56, 924–929. [PubMed: 27135806]
26. Marrink SJ; Berendsen HJC, Permeation Process of Small Molecules Across Lipid Membranes Studied by Molecular Dynamics Simulations. *J. Phys. Chem* 1996, 100, 16729–16738.
27. Swift RV; Amaro RE, Back to the Future: Can Physical Models of Passive Membrane Permeability Help Reduce Drug Candidate Attrition and Move Us Beyond QSPR? *Chemical Biology & Drug Design* 2013, 81, 61–71. [PubMed: 23066853]
28. Ghysels A; Venable RM; Pastor RW; Hummer G, Position-Dependent Diffusion Tensors in Anisotropic Media from Simulation: Oxygen Transport in and through Membranes. *J. Chem. Theory Comput* 2017, 13, 2962–2976. [PubMed: 28482659]
29. Comer J; Schulten K; Chipot C, Calculation of Lipid-Bilayer Permeabilities Using an Average Force. *J. Chem. Theory Comput* 2014, 10, 554–564. [PubMed: 26580032]
30. Lee CT; Comer J; Herndon C; Leung N; Pavlova A; Swift RV; Tung C; Rowley CN; Amaro RE; Chipot C; Wang Y; Gumbart JC, Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds. *J. Chem. Inf. Model* 2016, 56, 721–733. [PubMed: 27043429]
31. Mori T; Miyashita N; Im W; Feig M; Sugita Y, Molecular dynamics simulations of biological membranes and membrane proteins using enhanced conformational sampling algorithms. *BBA Biomemb* 2016, 1858, 1635–1651.
32. Still WC; Tempczyk A; Hawley RC; Hendrickson T, Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc* 1990, 112, 6127–6129.
33. Marrink SJ; Berendsen HJC, Simulation of Water Transport through a Lipid-Membrane. *J. Phys. Chem* 1994, 98, 4155–4168.
34. Mirjalili V; Feig M, Interactions of Amino Acid Side-Chain Analogs within Membrane Environments. *J. Phys. Chem. B* 2015, 119, 2877–2885. [PubMed: 25621811]
35. Tanizaki S; Feig M, A Generalized Born Formalism for Heterogeneous Dielectric Environments: Application to the Implicit Modeling of Biological Membranes. *J. Chem. Phys* 2005, 122, 124706. [PubMed: 15836408]
36. Sitkoff D; Sharp KA; Honig B, Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem* 1994, 98, 1978–1988.
37. Panahi A; Feig M, Dynamic Heterogeneous Dielectric Generalized Born (DHDGB): An Implicit Membrane Model with a Dynamically Varying Bilayer Thickness. *J. Chem. Theory Comput* 2013, 9, 1709–1719. [PubMed: 23585740]
38. Helfrich W, Elastic Properties of Lipid Bilayers - Theory and Possible Experiments. *Zeitschrift Fur Naturforschung C-a Journal of Biosciences* 1973, C 28, 693–703.
39. Xiang TX; Anderson BD, The Relationship between Permeant Size and Permeability in Lipid Bilayer-Membranes. *Journal of Membrane Biology* 1994, 140, 111–122. [PubMed: 7932645]

40. Korson L; Drosthan W; Millero FJ, Viscosity of Water at Various Temperatures. *J. Phys. Chem* 1969, 73, 34–39.
41. Brooks BR; Bruccoleri RE; Olafson BD; States DJ; Swaminathan S; Karplus M, CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem* 1983, 4, 187–217.
42. de la Torre JG; Huertas ML; Carrasco B, Calculation of Hydrodynamic Properties of Globular Proteins from their Atomic-Level Structure. *Biophys. J* 2000, 78, 719–730. [PubMed: 10653785]
43. Babic S; Horvat AJM; Pavlovic DM; Kastelan-Macan M, Determination of pK(a) Values of Active Pharmaceutical Ingredients. *TRAC Trends in Analytical Chemistry* 2007, 26, 1043–1061.
44. Barlow RB; Chan M, The Effects of Ph on the Affinity of Pirenzepine for Muscarinic Receptors in the Guinea-Pig Ileum and Rat Fundus Strip. *British Journal of Pharmacology* 1982, 77, 559–563. [PubMed: 6897199]
45. Cazares-Delgadillo J; Balaguer-Fernandez C; Calatayud-Pascual A; Ganem-Rondero A; Quintanar-Guerrero D; Lopez-Castellano AC; Merino V; Kalia YN, Transdermal Iontophoresis of Dexamethasone Sodium Phosphate in Vitro and in Vivo: Effect of Experimental Parameters and Skin Type on Drug Stability and Transport Kinetics. *European Journal of Pharmaceutics and Biopharmaceutics* 2010, 75, 173–178. [PubMed: 20332024]
46. Chamberlain K; Evans AA; Bromilow RH, 1-Octanol/Water Partition Coefficient (Kow) and pKa for Ionisable Pesticides Measured by apH-metric Method. *Pesticide Science* 1996, 47, 265–271.
47. Hurwitz AR; Liu ST, Determination of Aqueous Solubility and pKa Values of Estrogens. *J. Pharm. Sci* 1977, 66, 624–627. [PubMed: 17710]
48. Irwin JJ; Sterling T; Mysinger MM; Bolstad ES; Coleman RG, ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model* 2012, 52, 1757–1768. [PubMed: 22587354]
49. Kim S; Thiessen PA; Bolton EE; Chen J; Fu G; Gindulyte A; Han LY; He JE; He SQ; Shoemaker BA; Wang JY; Yu B; Zhang J; Bryant SH, PubChem Substance and Compound Databases. *Nucleic Acids Res* 2016, 44, D1202–D1213. [PubMed: 26400175]
50. Kortejarvi H; Yliperttula M; Dressman JB; Junginger HE; Midha KK; Shah VP; Barends DM, Biowaiver Monographs for Immediate Release Solid Oral Dosage Forms: Ranitidine Hydrochloride. *J. Pharm. Sci* 2005, 94, 1617–1625. [PubMed: 15959881]
51. Martinez V; Maguregui MI; Jimenez RM; Alonso RM, Determination of the pKa Values of Beta-Blockers by Automated Potentiometric Titrations. *J. Pharm. Biomed. Anal* 2000, 23, 459–468. [PubMed: 10933539]
52. Oo CY; Kuhn RJ; Desai N; Wright CE; Mcnamara PJ, Pharmacokinetics in Lactating Women - Prediction of Alprazolam Transfer into Milk. *British Journal of Clinical Pharmacology* 1995, 40, 231–236. [PubMed: 8527284]
53. Prankerd RJ, Critical Compilation of pKa Values for Pharmaceutical Substances. *Profiles of Drug Substances, Excipients and Related Methodology* 2007, 33, 1–33.
54. Rodriguez-Barrientos D; Rojas-Hernandez A; Gutierrez A; Moya-Hernandez R; Gomez-Balderas R; Ramirez-Silva MT, Determination of pKa values of Tenoxicam from <sup>1</sup>H NMR Chemical Shifts and of Oxicams from Electrophoretic Mobilities (CZE) with the aid of Programs SQUAD and HYPNMR. *Talanta* 2009, 80, 754–762. [PubMed: 19836548]
55. Shalaeva M; Kenseth J; Lombardo F; Bastinz A, Measurement of Dissociation Constants (pK(a) Values) of Organic Compounds by Multiplexed Capillary Electrophoresis using Aqueous and Cosolvent Buffers. *J. Pharm. Sci* 2008, 97, 2581–2606. [PubMed: 18228610]
56. Wishart DS; Knox C; Guo AC; Shrivastava S; Hassanali M; Stothard P; Chang Z; Woolsey J, DrugBank: a Comprehensive Resource for in Silico Drug Discovery and Exploration. *Nucleic Acids Res* 2006, 34, D668–D672. [PubMed: 16381955]
57. Lide DR, *Handbook of Chemistry and Physics*, 84th Edition CRC Press: Boca Raton, 2003.
58. Moffat A; Osselton D; Widdop B; Watts J, Clarke's Analysis of Drugs and Poisons Pharmaceutical Press: London, 2011.
59. Filaretova LP; Takeuchi K, *Cell/Tissue Injury and Cytoprotection/Organoprotection in the Gastrointestinal Tract*. Karger Medical and Scientific Publishers: Basel, 2012.

60. Sourì E; Farahani AD; Ahmadvkhaniha R; Amini M, A Stability Indicating HPLC Method for the Determination of Clobazam and its Basic Degradation Product Characterization. Daru, Journal of Pharmaceutical Sciences 2014, 22, 49. [PubMed: 24919821]
61. Committee JFWE, Toxicological Evaluation of Certain Veterinary Drug Residues in Food World Health Organization: Geneva, 2005.
62. Connors KA; Amidon GL; Stella VJ, Chemical Stability of Pharmaceutical: A Handbook for Pharmacists John Wiley & Sons: Hoboken, 1986.
63. Bolton EE; Chen J; Kim S; Han LY; He SQ; Shi WY; Simonyan V; Sun Y; Thiessen PA; Wang JY; Yu B; Zhang J; Bryant SH, PubChem3D: A New Resource for Scientists. J. Cheminf 2011, 3.
64. Feig M; Karanicolas J; Brooks CL, MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. Journal of Molecular Graphics and Modelling 2004, 22, 377–395. [PubMed: 15099834]
65. Vanommeslaeghe K; Hatcher E; Acharya C; Kundu S; Zhong S; Shim J; Darian E; Guvench O; Lopes P; Vorobyov I; MacKerell AD, CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem 2010, 31, 671–690. [PubMed: 19575467]
66. Vanommeslaeghe K; MacKerell AD, Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. J. Chem. Inf. Model 2012, 52, 3144–3154. [PubMed: 23146088]
67. Vanommeslaeghe K; Raman EP; MacKerell AD, Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. J. Chem. Inf. Model 2012, 52, 3155–3168. [PubMed: 23145473]
68. Case DA; Betz RM; Botello-Smith W; Cerutti DS; Cheatham TE; Darden TA; Duke RE; Giese TJ; Gohlke H; Goetz AW; Homeyer N; Izadi S; Janowski P; Kaus J; Kovalenko A; Lee TS; LeGrand S; Li P; Lin C; Luchko T; Luo R; Madej B; Mermelstein D; Merz KM; Monard G; Nguyen H; Nguyen HT; Omelyan I; Onufriev A; Roe DR; Roitberg A; Sagui C; Simmerling CL; Swails J; Walker RC; Wang J; Wolf RM; Wu X; Xiao L; York DM; Kollman PA Amber 2016, University of California: San Francisco: 2016.
69. Jakalian A; Bush BL; Jack DB; Bayly CI, Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. J. Comput. Chem 2000, 21, 132–146.
70. Jakalian A; Jack DB; Bayly CI, Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. J. Comput. Chem 2002, 23, 1623–1641. [PubMed: 12395429]
71. Banks JL; Beard HS; Cao YX; Cho AE; Damm W; Farid R; Felts AK; Halgren TA; Mainz DT; Maple JR; Murphy R; Philipp DM; Repasky MP; Zhang LY; Berne BJ; Friesner RA; Gallicchio E; Levy RM, Integrated Modeling Program, Applied Chemical Theory (IMPACT). J. Comput. Chem 2005, 26, 1752–1780. [PubMed: 16211539]
72. Jorgensen WL; Maxwell DS; TiradoRives J, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J. Am. Chem. Soc 1996, 118, 11225–11236.
73. Schrödinger Release 2018-1: LigPrep, Schrödinger, LLC New York, NY, 2018.
74. Bondi A, Van der Waals Volumes + Radii. J. Phys. Chem 1964, 68, 441–451.
75. Abadi M; Agarwal A; Barham P; Brevdo E; Chen Z; Citro C; Corrado GS; Davis A; Dean J; Devin M; Ghemawat S; Goodfellow I; Harp A; Irving G; Isard M; Jia Y; Jozefowicz R; Kaiser L; Kudlur M; Levenberg J; Mane D; Monga R; Moore S; Murray D; Olah C; Schuster M; Shlens J; Steiner B; Sutskever I; Talwar K; Tucker P; Vanhoucke V; Vasudevan V; Viegas F; Vinyals O; Warden P; Wattenberg M; Wicke M; Yu Y; Zheng X, In; arXiv: 2016.
76. Cheng H-T; Koc L; Harmsen J; Shaked T; Chandra T; Aradhye H; Anderson G; Corrado G; Chai W; Ispir M; Anil R; Haque Z; Hong L; Jain V; Liu X; Shah H, In; arXiv: 2016.
77. Seibold S; Perktold J, In 9th Python in Science Conference; 2010.
78. Avdeef A; Artursson P; Neuhoff S; Lazorova L; Grasjo J; Tavelin S, Caco-2 Permeability of Weakly Basic Drugs Predicted with the Double-Sink PAMPA pKa(flux) Method. European Journal of Pharmaceutical Sciences 2005, 24, 333–349. [PubMed: 15734300]

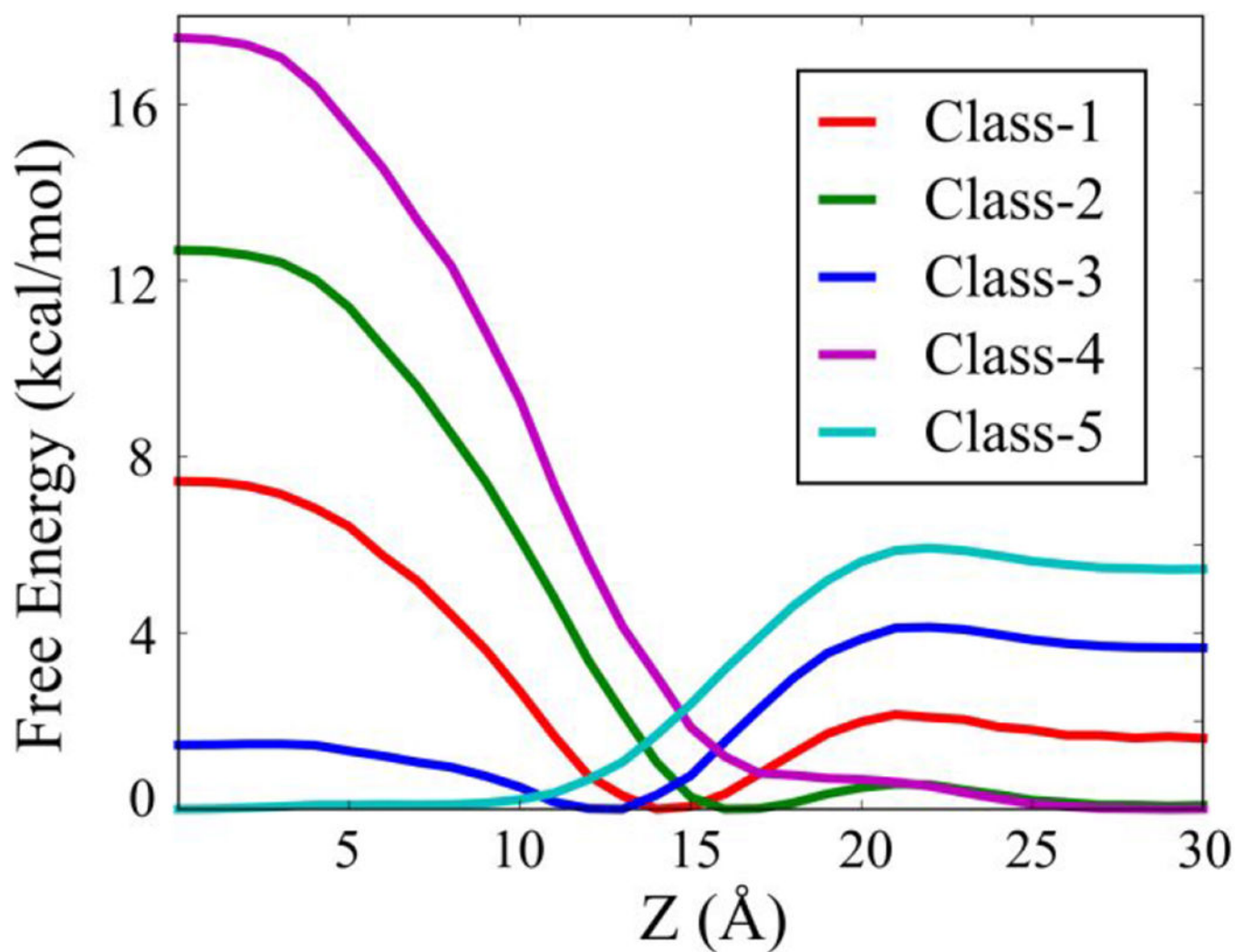
79. Balimane PV; Han YH; Chong SH, Current Industrial Practices of Assessing Permeability and P-Glycoprotein Interaction. *Aaps Journal* 2006, 8, E1–E13. [PubMed: 16584115]
80. Bermejo M; Avdeef A; Ruiz A; Nalda R; Ruell JA; Tsinman O; Gonzalez I; Fernandez C; Sanchez G; Garrigues TM; Merino V, PAMPA - a Drug Absorption in Vitro Model 7. Comparing Rat in situ, Caco-2, and PAMPA Permeability of Fluoroquinolones. *European Journal of Pharmaceutical Sciences* 2004, 21, 429–441. [PubMed: 14998573]
81. Di L; Kerns EH; Fan K; McConnell OJ; Carter GT, High Throughput Artificial Membrane Permeability Assay for Blood-Brain Barrier. *European Journal of Medicinal Chemistry* 2003, 38, 223–232. [PubMed: 12667689]
82. Fujikawa M; Nakao K; Shimizu R; Akamatsu M, QSAR Study on Permeability of Hydrophobic Compounds with Artificial Membranes. *Bioorganic & Medicinal Chemistry* 2007, 15, 3756–3767. [PubMed: 17418579]
83. Carrie I; Clement M; de Javel D; Frances H; Bourre JM, Specific Phospholipid Fatty Acid Composition of Brain Regions in Mice: Effects of n-3 Polyunsaturated Fatty Acid Deficiency and Phospholipid Supplementation. *Journal of Lipid Research* 2000, 41, 465–472. [PubMed: 10706594]
84. Schwarz SM; Ling S; Hostetler B; Draper JP; Watkins JB, Lipid-Composition and Membrane Fluidity in the Small-Intestine of the Developing Rabbit. *Gastroenterology* 1984, 86, 1544–1551. [PubMed: 6714577]
85. Zhou MM; Xue Y; Sun SH; Wen M; Li ZJ; Xu J; Wang JF; Yanagita T; Wang YM; Xue CH, Effects of Different Fatty Acids Composition of Phosphatidylcholine on Brain Function of Dementia Mice Induced by Scopolamine. *Lipids in Health and Disease* 2016, 15, 135. [PubMed: 27558491]
86. Nagle JF; Tristram-Nagle S, Structure of Lipid Bilayers. *Biochim. Biophys. Acta* 2000, 1469, 159–195. [PubMed: 11063882]
87. Avdeef A; Tsinman O, PAMPA - A drug Absorption in Vitro Model 13. Chemical Selectivity due to Membrane Hydrogen Bonding: In Combo Comparisons of HDM-, DOPC-, and DS-PAMPA Models. *European Journal of Pharmaceutical Sciences* 2006, 28, 43–50. [PubMed: 16476536]
88. Yamashita F; Wanchana S; Hashida M, Quantitative Structure/Property Relationship Analysis of Caco-2 Permeability Using a Genetic Algorithm-Based Partial Least Squares Method. *J. Pharm. Sci* 2002, 91, 2230–2239. [PubMed: 12226850]



**Figure 1.**

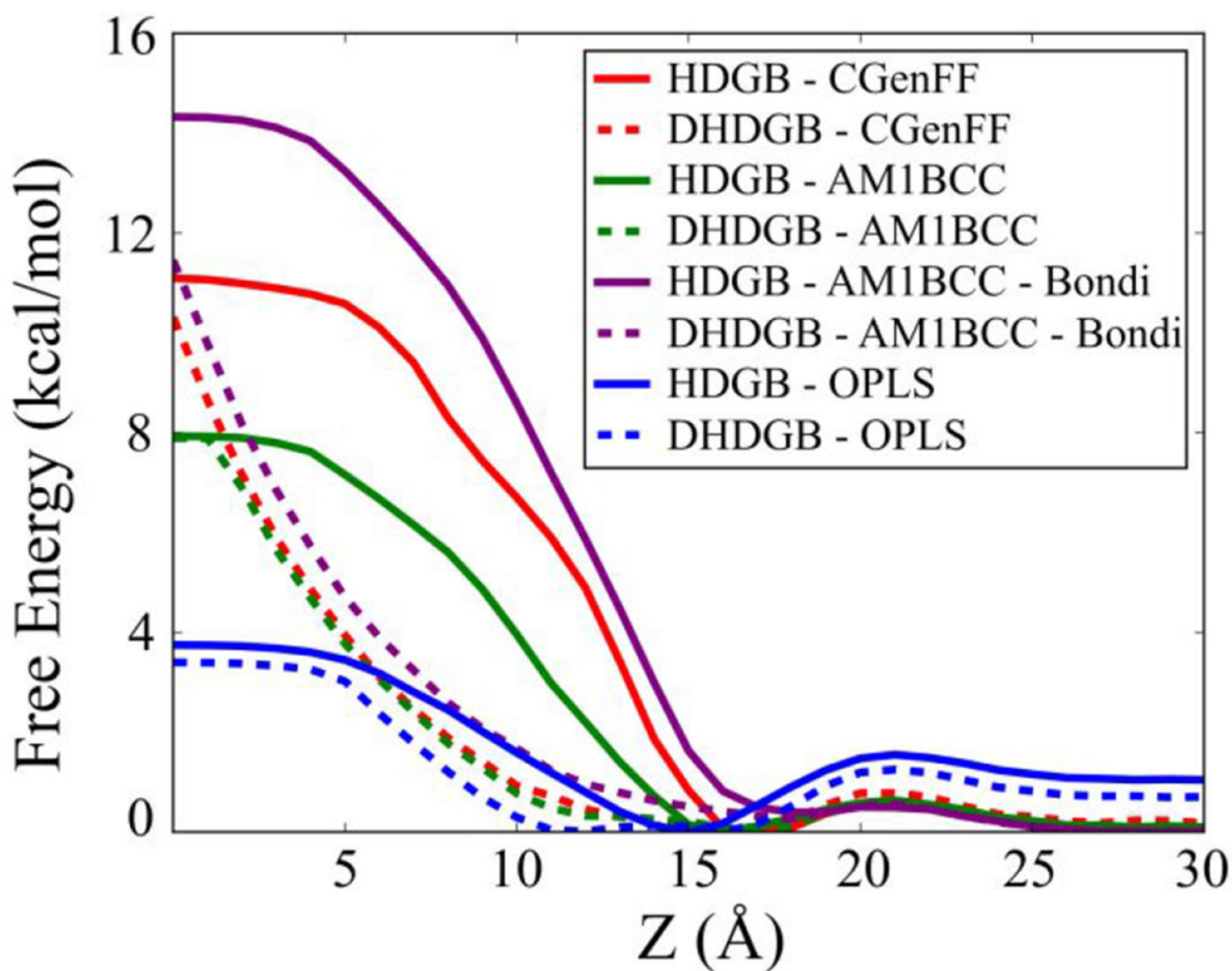
Alternate estimates of  $G_{tr}$  (A) and  $\delta_{barrier}$  (B) from insertion free energy profiles along the membrane normal. Energy values at specific points of the curve are referred to as ‘ $E$ ’ values and areas under the curve are ‘ $I$ ’ quantities.  $E_{max}$  is the maximum energy,  $E_{min}$  is the minimum energy,  $E_{avg}$  is the average energy over a range of  $z$  positions,  $E_{locmax}$  is the energy at the local maximum upon reaching the membrane surface from the position where  $E_{min}$  is observed, and  $E_{locmin}$  is the energy at the local minimum between the position of  $E_{locmax}$  and bulk water. If any of these barriers is not observed, then  $\delta_{barrier}$  was set to a small value chosen as 1 Å since it is the resolution of our free energy profiles.



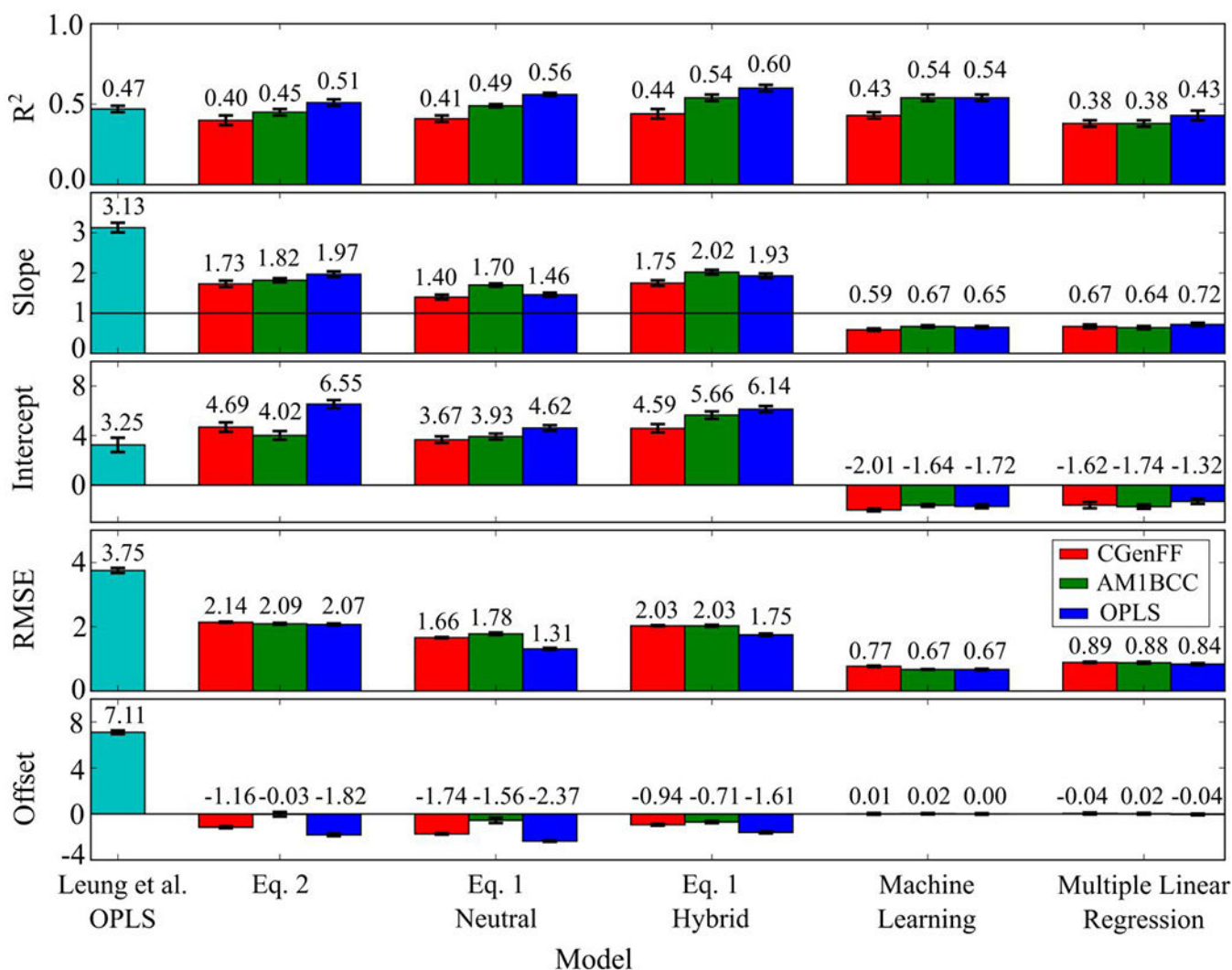


**Figure 2.**  
Classification of profiles of free energy of membrane insertion as described in Table 1.

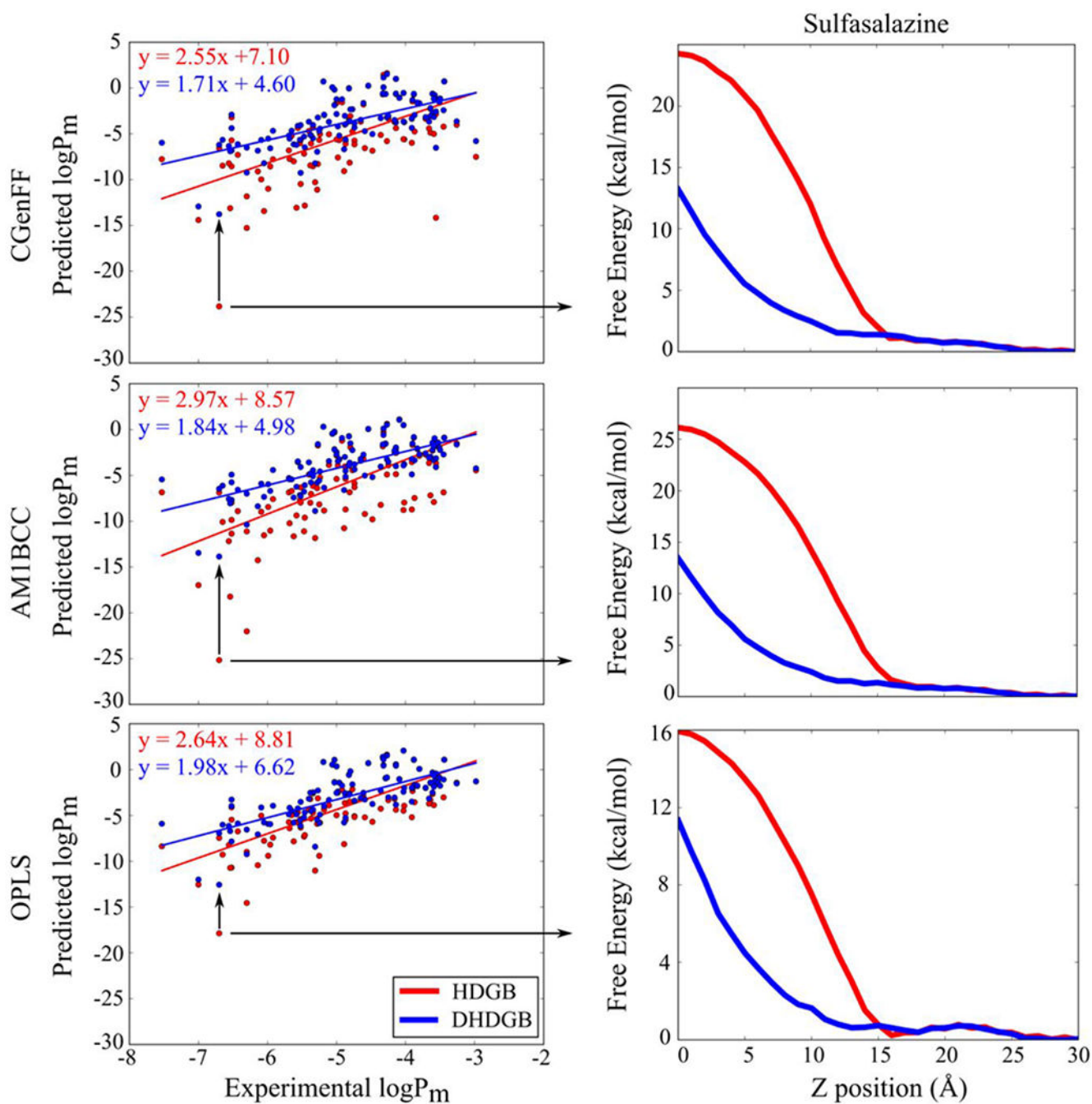




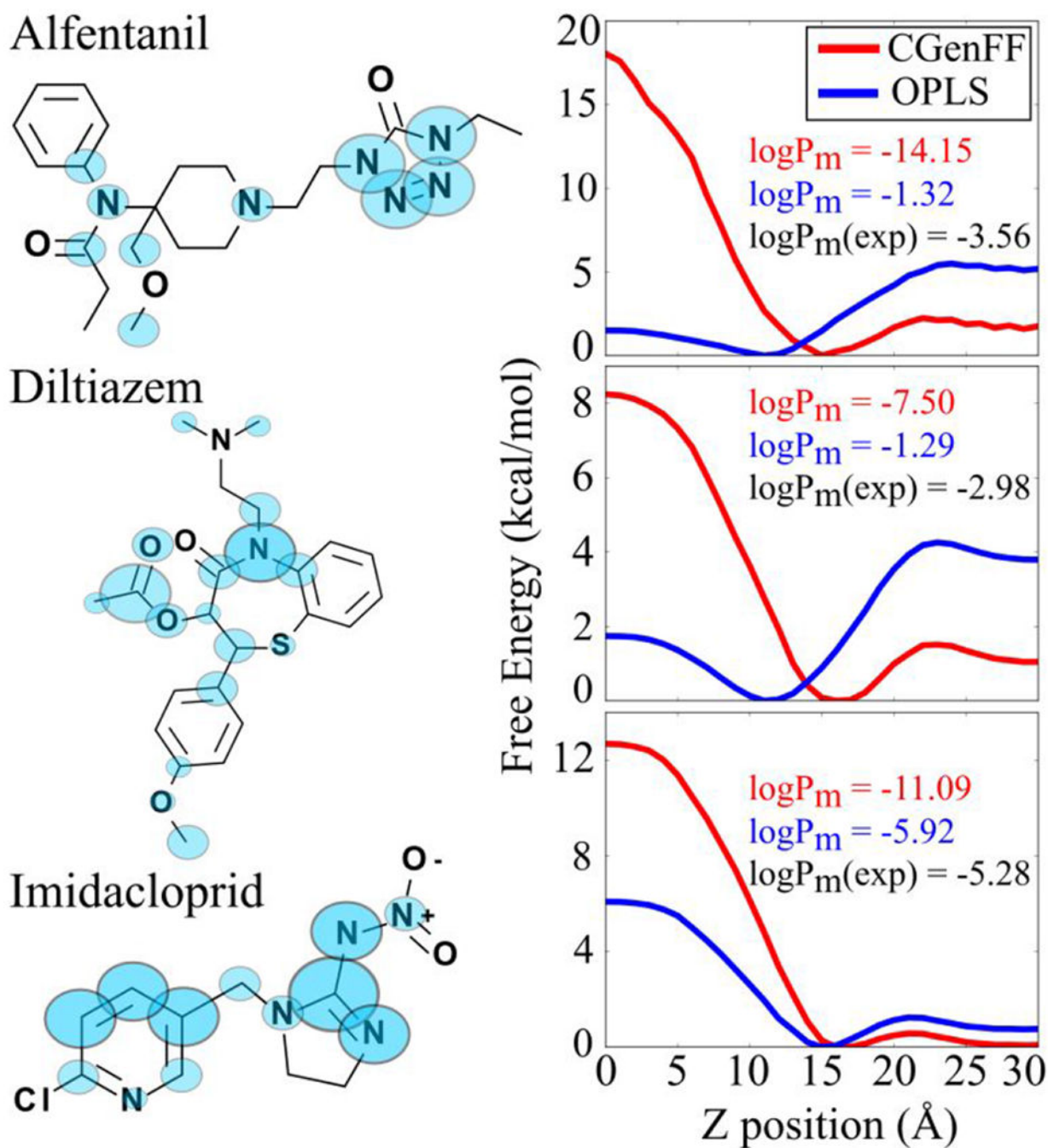
**Figure 3.** Representative membrane insertion profiles for caffeine with HDGB (solid lines) and DHDGB (dashed lines) with atomic radii from the CHARMM Lennard-Jones parameters and CGenFF (red), AM1-BCC (green), and OPLS (blue) charge sets as well as Bondi radii and AM1-BCC charges (purple).

**Figure 4.**

Summary of the overall evaluation of models with different charge sets. DHDGB profiles were used for Eq. 1 and 2 predictions and for machine learning and multiple linear regression models.  $R^2$ , slope, intercept, RMSE, and offset values as in Table 2.



**Figure 5.** Correlation plots of predicted  $\log P_m$  values for HDGB and DHDGB models calculated by Eq. 2 and used with partial charges from CGenFF, AM1-BCC and OPLS. The linear regression equations are given for HDGB (red) and DHDGB (blue) inside the plot frames. At the right, the free energy profiles of sulfasalazine are shown.



**Figure 6.** Structures of alfentanil, diltiazem and imidacloprid and differences in their free energy profiles from CGenFF and OPLS charges obtained by HDGB. The circles in the structures illustrate the differences in the partial charges between CGenFF and OPLS; larger circles mean larger differences.

**Table 1.**

Classification criteria of free energy profiles.

Class	Criterion
1	$E_{30} < E_0$ ; $E_{30} > 1.0$ kcal/mol or $E_{30} / E_0 > 0.1$
2	$E_{30} < E_0$ ; $E_0 < 1.0$ kcal/mol; $E_{30} / E_0 < 0.1$ ; $z_{min} < 25$ Å.
3	$E_{30} > E_0$ ; $E_0 > 0.2$ kcal/mol
4	$E_{30} < E_0$ ; $E_{30} < 1.0$ kcal/mol; $E_{30} / E_0 < 0.1$ ; $z_{min} > 25$ Å.
5	$E_{30} > E_0$ ; $E_0 < 0.2$ kcal/mol

$E_{30}$  is the energy in the water phase at  $z=30$ ,  $E_0$  is the energy in the bilayer at  $z=0$  and  $z_{min}$  is the  $z$ -position where minimum energy is obtained.

Table 2.

Evaluation of the  $\log P_m$  predictions for the neutral free energy profiles based on integration of the membrane insertion and Eq. 1.

	HDGB				DHDGB					
	$R^2$	S	I	RMSE	Offset	$R^2$	S	I	RMSE	Offset
Water Diffusion Profile										
CGenFF	0.38 ±0.02	1.69 ±0.09	6.61 ±0.36	2.22 ±0.14	-3.27 ±0.12	0.41 ±0.02	1.38 ±0.06	5.58 ±0.27	1.63 ±0.02	-3.74 ±0.05
AM1-BCC	0.45 ±0.01	2.34 ±0.13	8.10 ±0.44	2.85 ±0.20	-1.60 ±0.27	0.50 ±0.01	1.69 ±0.04	5.91 ±0.23	1.76 ±0.05	-2.56 ±0.23
AM1-BCC-Bondi	0.44 ±0.02	2.63 ±0.22	8.23 ±0.79	3.32 ±0.32	-0.33 ±0.31	0.46 ±0.02	1.75 ±0.05	5.21 ±0.26	1.93 ±0.03	-1.59 ±0.07
OPLS	0.55 ±0.01	1.65 ±0.07	7.23 ±0.25	1.58 ±0.08	-4.08 ±0.08	0.56 ±0.01	1.42 ±0.05	6.42 ±0.22	1.27 ±0.03	-4.37 ±0.05
Diffusion Profile from Bemporad <i>et al.</i>										
CGenFF	0.37 ±0.02	1.71 ±0.10	6.02 ±0.37	2.26 ±0.14	-2.56 ±0.12	0.40 ±0.02	1.39 ±0.06	4.97 ±0.27	1.66 ±0.02	-3.05 ±0.05
AM1-BCC	0.45 ±0.01	2.33 ±0.13	7.36 ±0.44	2.86 ±0.20	-0.88 ±0.27	0.49 ±0.01	1.67 ±0.04	5.15 ±0.24	1.76 ±0.04	-1.88 ±0.23
AM1-BCC-Bondi	0.43 ±0.02	2.60 ±0.22	7.38 ±0.78	3.30 ±0.32	0.40 ±0.31	0.46 ±0.02	1.71 ±0.05	4.37 ±0.25	1.91 ±0.03	-0.90 ±0.07
OPLS	0.55 ±0.01	1.71 ±0.07	6.88 ±0.25	1.65 ±0.08	-3.41 ±0.09	0.56 ±0.02	1.48 ±0.05	6.06 ±0.23	1.33 ±0.03	-3.74 ±0.05
Diffusion Profile from Carpenter <i>et al.</i>										
CGenFF	0.38 ±0.02	1.70 ±0.10	6.07 ±0.37	2.24 ±0.14	-2.67 ±0.12	0.41 ±0.02	1.38 ±0.06	5.02 ±0.27	1.63 ±0.02	-3.15 ±0.05
AM1-BCC	0.45 ±0.01	2.33 ±0.13	7.47 ±0.44	2.85 ±0.20	-1.01 ±0.27	0.49 ±0.01	1.67 ±0.04	5.26 ±0.24	1.75 ±0.04	-2.00 ±0.23
AM1-BCC-Bondi	0.43 ±0.02	2.61 ±0.22	7.53 ±0.79	3.30 ±0.32	0.27 ±0.31	0.46 ±0.02	1.72 ±0.05	4.51 ±0.26	1.91 ±0.03	-1.03 ±0.07
OPLS	0.55 ±0.01	1.69 ±0.07	6.85 ±0.25	1.62 ±0.08	-3.50 ±0.08	0.56 ±0.02	1.45 ±0.05	6.03 ±0.23	1.30 ±0.03	-3.82 ±0.05
Diffusion Profile with a Higher Viscosity for Membrane.										
CGenFF	0.38 ±0.02	1.71 ±0.09	4.69 ±0.36	2.25 ±0.14	-1.26 ±0.12	0.41 ±0.02	1.40 ±0.06	3.67 ±0.27	1.66 ±0.02	-1.74 ±0.05
AM1-BCC	0.45 ±0.01	2.35 ±0.13	6.12 ±0.44	2.87 ±0.20	0.41 ±0.27	0.49 ±0.01	1.70 ±0.04	3.93 ±0.24	1.78 ±0.04	-0.56 ±0.23
AM1-BCC-Bondi	0.43 ±0.02	2.63 ±0.22	6.21 ±0.79	3.33 ±0.32	1.70 ±0.31	0.46 ±0.02	1.75 ±0.05	3.19 ±0.26	1.94 ±0.03	0.43 ±0.07
OPLS	0.55 ±0.01	1.68 ±0.06	5.41 ±0.25	1.61 ±0.08	-2.08 ±0.08	0.56 ±0.01	1.46 ±0.05	4.62 ±0.23	1.31 ±0.03	-2.37 ±0.05

$R^2$ , slope (S) and intercept (I) were calculated from the linear regression curves between predictions and experimental values. RMSE were calculated against the experimental values using an offset for each case to find the minimum possible RMSE values. Average values and standard errors were calculated among subsets of the drugs which served as the test sets in the machine learning method.



Table 3.

Evaluation of the  $\log P_m$  predictions for hybrid membrane insertion profiles obtained from charged and neutral free energy profiles and by the addition of  $G_{state}$  values calculated from experimental  $pK_a$  values.

	HDGB				DHDGB					
	$R^2$	S	I	RMSE	Offset	$R^2$	S	I	RMSE	Offset
Water Diffusion Profile										
CGenFF	0.42 ±0.02	2.11 ±0.09	7.71 ±0.37	2.63 ±0.16	-2.33 ±0.13	0.47 ±0.02	1.76 ±0.05	6.54 ±0.28	1.95 ±0.02	-2.84 ±0.06
AM1-BCC	0.47 ±0.02	2.54 ±0.13	9.32 ±0.48	3.02 ±0.21	-1.86 ±0.19	0.52 ±0.02	2.06 ±0.06	7.74 ±0.28	2.14 ±0.02	-2.59 ±0.09
AM1-BCC-Bondi	0.42 ±0.01	2.95 ±0.23	8.83 ±0.85	3.88 ±0.35	0.65 ±0.32	0.44 ±0.02	2.02 ±0.07	5.63 ±0.38	2.38 ±0.04	-0.66 ±0.08
OPLS	0.54 ±0.01	2.15 ±0.06	8.73 ±0.24	2.21 ±0.10	-3.13 ±0.10	0.57 ±0.02	1.92 ±0.04	7.91 ±0.17	1.81 ±0.02	-3.46 ±0.07
Diffusion Profile from Bemporad <i>et al.</i>										
CGenFF	0.42 ±0.02	2.11 ±0.10	7.03 ±0.38	2.65 ±0.16	-1.64 ±0.14	0.46 ±0.02	1.76 ±0.06	5.83 ±0.28	1.96 ±0.03	-2.18 ±0.08
AM1-BCC	0.47 ±0.02	2.53 ±0.13	8.56 ±0.49	3.03 ±0.21	-1.12 ±0.20	0.52 ±0.02	2.04 ±0.06	6.94 ±0.28	2.13 ±0.03	-1.89 ±0.09
AM1-BCC-Bondi	0.42 ±0.01	2.93 ±0.23	7.97 ±0.84	3.85 ±0.35	1.40 ±0.33	0.44 ±0.02	1.99 ±0.07	4.78 ±0.37	2.34 ±0.04	0.04 ±0.08
OPLS	0.53 ±0.01	2.18 ±0.07	8.18 ±0.25	2.27 ±0.10	-2.44 ±0.10	0.57 ±0.02	1.93 ±0.04	7.32 ±0.19	1.84 ±0.02	-2.80 ±0.07
Diffusion Profile from Carpenter <i>et al.</i>										
CGenFF	0.42 ±0.02	2.11 ±0.10	7.11 ±0.38	2.63 ±0.16	-1.75 ±0.14	0.46 ±0.02	1.75 ±0.06	5.91 ±0.28	1.94 ±0.03	-2.28 ±0.08
AM1-BCC	0.47 ±0.02	2.53 ±0.13	8.67 ±0.49	3.02 ±0.21	-1.24 ±0.19	0.52 ±0.02	2.04 ±0.06	7.05 ±0.27	2.12 ±0.03	-1.99 ±0.09
AM1-BCC-Bondi	0.42 ±0.01	2.93 ±0.23	8.13 ±0.85	3.85 ±0.35	1.25 ±0.33	0.44 ±0.02	2.00 ±0.07	4.92 ±0.37	2.34 ±0.04	-0.10 ±0.08
OPLS	0.54 ±0.01	2.16 ±0.07	8.21 ±0.25	2.24 ±0.10	-2.55 ±0.10	0.57 ±0.02	1.92 ±0.04	7.35 ±0.18	1.82 ±0.02	-2.90 ±0.07
Diffusion Profile with a Higher Viscosity for Membrane.										
CGenFF	0.42 ±0.02	2.12 ±0.10	5.76 ±0.37	2.65 ±0.16	-0.32 ±0.13	0.46 ±0.02	1.78 ±0.06	4.60 ±0.28	1.97 ±0.03	-0.82 ±0.07
AM1-BCC	0.47 ±0.02	2.55 ±0.13	7.35 ±0.49	3.05 ±0.21	0.17 ±0.19	0.52 ±0.02	2.07 ±0.06	5.77 ±0.28	2.16 ±0.03	-0.55 ±0.09
AM1-BCC-Bondi	0.42 ±0.01	2.96 ±0.23	6.80 ±0.85	3.89 ±0.35	2.69 ±0.32	0.44 ±0.02	2.03 ±0.07	3.60 ±0.38	2.38 ±0.04	1.37 ±0.08
OPLS	0.54 ±0.01	2.17 ±0.06	6.83 ±0.24	2.24 ±0.10	-1.13 ±0.10	0.57 ±0.02	1.94 ±0.04	6.02 ±0.18	1.84 ±0.02	-1.46 ±0.07

$R^2$ , slope (S), intercept (I), RMSE, and offset values as in Table 2.



Evaluation of the  $\log P_m$  predictions for hybrid membrane insertion profiles obtained from charged and neutral free energy profiles and by the addition of  $G_{state}$  values predicted by Leung *et al.*

Table 4.

	HDGB				DHDGB					
	$R^2$	S	I	RMSE	Offset	$R^2$	S	I	RMSE	Offset
Water Diffusion Profile										
CGenFF	0.41 ±0.02	2.09 ±0.11	7.78 ±0.42	2.67 ±0.16	-2.48 ±0.13	0.44 ±0.03	1.73 ±0.07	6.53 ±0.34	2.00 ±0.02	-2.96 ±0.07
AM1-BCC	0.48 ±0.02	2.51 ±0.13	9.34 ±0.51	2.94 ±0.22	-2.00 ±0.19	0.54 ±0.02	2.01 ±0.06	7.64 ±0.30	2.01 ±0.03	-2.71 ±0.09
AM1-BCC-Bondi	0.43 ±0.02	2.94 ±0.23	8.90 ±0.87	3.82 ±0.35	0.54 ±0.32	0.45 ±0.02	2.00 ±0.08	5.60 ±0.40	2.31 ±0.04	-0.74 ±0.09
OPLS	0.56 ±0.02	2.15 ±0.07	8.91 ±0.28	2.16 ±0.11	-3.29 ±0.10	0.60 ±0.02	1.91 ±0.06	8.03 ±0.25	1.72 ±0.04	-3.61 ±0.08
Diffusion Profile from Bemporad <i>et al.</i>										
CGenFF	0.40 ±0.02	2.10 ±0.11	7.11 ±0.43	2.70 ±0.16	-1.75 ±0.14	0.43 ±0.03	1.73 ±0.07	5.83 ±0.34	2.01 ±0.02	-2.29 ±0.08
AM1-BCC	0.48 ±0.02	2.50 ±0.14	8.58 ±0.51	2.94 ±0.22	-1.27 ±0.19	0.54 ±0.02	1.99 ±0.06	6.84 ±0.30	2.00 ±0.03	-2.01 ±0.09
AM1-BCC-Bondi	0.43 ±0.02	2.92 ±0.24	8.05 ±0.87	3.78 ±0.35	1.27 ±0.33	0.45 ±0.02	1.97 ±0.08	4.75 ±0.39	2.27 ±0.04	-0.06 ±0.09
OPLS	0.55 ±0.02	2.18 ±0.08	8.35 ±0.29	2.21 ±0.11	-2.59 ±0.11	0.59 ±0.02	1.92 ±0.06	7.43 ±0.25	1.75 ±0.03	-2.94 ±0.08
Diffusion Profile from Carpenter <i>et al.</i>										
CGenFF	0.41 ±0.02	2.09 ±0.11	7.18 ±0.43	2.67 ±0.16	-1.87 ±0.14	0.44 ±0.03	1.72 ±0.07	5.90 ±0.34	1.99 ±0.02	-2.39 ±0.08
AM1-BCC	0.48 ±0.02	2.50 ±0.14	8.68 ±0.51	2.93 ±0.22	-1.40 ±0.19	0.54 ±0.02	1.99 ±0.06	6.95 ±0.30	1.99 ±0.03	-2.13 ±0.09
AM1-BCC-Bondi	0.43 ±0.02	2.92 ±0.24	8.20 ±0.87	3.79 ±0.35	1.15 ±0.33	0.45 ±0.02	1.97 ±0.08	4.89 ±0.39	2.27 ±0.04	-0.18 ±0.09
OPLS	0.56 ±0.02	2.17 ±0.08	8.38 ±0.29	2.18 ±0.11	-2.69 ±0.11	0.60 ±0.02	1.91 ±0.06	7.46 ±0.25	1.73 ±0.04	-3.04 ±0.08
Diffusion Profile with a Higher Viscosity for Membrane.										
CGenFF	0.41 ±0.02	2.11 ±0.11	5.83 ±0.42	2.70 ±0.16	-0.45 ±0.14	0.44 ±0.03	1.75 ±0.07	4.59 ±0.34	2.03 ±0.02	-0.94 ±0.08
AM1-BCC	0.48 ±0.02	2.52 ±0.13	7.37 ±0.51	2.96 ±0.22	0.01 ±0.19	0.54 ±0.02	2.02 ±0.06	5.66 ±0.30	2.03 ±0.03	-0.71 ±0.09
AM1-BCC-Bondi	0.43 ±0.02	2.94 ±0.23	6.88 ±0.87	3.82 ±0.35	2.56 ±0.32	0.45 ±0.02	2.00 ±0.08	3.57 ±0.40	2.32 ±0.04	1.28 ±0.09
OPLS	0.56 ±0.02	2.18 ±0.07	7.02 ±0.28	2.19 ±0.11	-1.29 ±0.10	0.60 ±0.02	1.93 ±0.06	6.14 ±0.25	1.75 ±0.04	-1.61 ±0.08

$R^2$ , slope (S), intercept (I), RMSE, and offset values as in Table 2.

**Table 5.**

Evaluation of the  $\log P_m$  predictions based on barrier approximation using Eq. 2 in comparison with values reported by Leung *et al.*<sup>24</sup> for the same drug sets.

	$R^2$	S	I	RMSE	Offset
HDGB					
CGenFF	0.39 $\pm$ 0.02	2.18 $\pm$ 0.13	6.17 $\pm$ 0.50	2.87 $\pm$ 0.17	-0.41 $\pm$ 0.19
AM1-BCC	0.43 $\pm$ 0.01	2.68 $\pm$ 0.16	6.89 $\pm$ 0.55	3.38 $\pm$ 0.22	1.28 $\pm$ 0.31
AM1-BCC-Bondi	0.42 $\pm$ 0.02	2.89 $\pm$ 0.28	6.65 $\pm$ 0.97	3.70 $\pm$ 0.37	2.54 $\pm$ 0.41
OPLS	0.50 $\pm$ 0.01	2.32 $\pm$ 0.10	7.74 $\pm$ 0.37	2.54 $\pm$ 0.11	-1.32 $\pm$ 0.14
DHDGB					
CGenFF	0.40 $\pm$ 0.03	1.73 $\pm$ 0.08	4.69 $\pm$ 0.38	2.14 $\pm$ 0.02	-1.16 $\pm$ 0.08
AM1-BCC	0.45 $\pm$ 0.02	1.82 $\pm$ 0.05	4.02 $\pm$ 0.35	2.09 $\pm$ 0.03	-0.03 $\pm$ 0.20
AM1-BCC-Bondi	0.42 $\pm$ 0.02	1.76 $\pm$ 0.08	2.91 $\pm$ 0.37	2.12 $\pm$ 0.04	0.82 $\pm$ 0.08
OPLS	0.51 $\pm$ 0.02	1.97 $\pm$ 0.07	6.55 $\pm$ 0.32	2.07 $\pm$ 0.03	-1.82 $\pm$ 0.11
Two-State (Leung <i>et al.</i> )					
OPLS	0.47 $\pm$ 0.02	3.13 $\pm$ 0.12	3.25 $\pm$ 0.58	3.75 $\pm$ 0.08	7.11 $\pm$ 0.16

Based on  $G_{\text{barrier}} = G_1$   $\delta_{\text{barrier}} = \delta_2$  and including  $G_{\text{state}}$  values given by Leung *et al.*;  $R^2$ , slope (S), intercept (I), RMSE, and offset values as in Table 2.

**Table 6.**Evaluation of  $\log P_m$  predictions for different insertion profiles widths with DHDGB and OPLS charges

Profile width change [ $\text{\AA}$ ]	$R^2$	Slope	Intercept	RMSE	Offset
Neutral Profile					
-1	$0.55 \pm 0.02$	$1.25 \pm 0.04$	$3.87 \pm 0.20$	$1.10 \pm 0.02$	$-2.64 \pm 0.04$
0	$0.56 \pm 0.01$	$1.46 \pm 0.05$	$4.62 \pm 0.23$	$1.31 \pm 0.03$	$-2.37 \pm 0.05$
+1	$0.57 \pm 0.01$	$1.53 \pm 0.05$	$4.83 \pm 0.23$	$1.37 \pm 0.03$	$-2.25 \pm 0.06$
Hybrid Profile ( $G_{state}$ from experimental $pK_a$ )					
-1	$0.57 \pm 0.02$	$1.71 \pm 0.03$	$5.27 \pm 0.15$	$1.58 \pm 0.02$	$-1.81 \pm 0.05$
0	$0.57 \pm 0.02$	$1.94 \pm 0.04$	$6.02 \pm 0.18$	$1.84 \pm 0.02$	$-1.46 \pm 0.07$
+1	$0.57 \pm 0.02$	$2.00 \pm 0.04$	$6.18 \pm 0.19$	$1.92 \pm 0.02$	$-1.31 \pm 0.08$
Hybrid Profile ( $G_{state}$ from Leung <i>et al.</i> )					
-1	$0.60 \pm 0.02$	$1.69 \pm 0.05$	$5.31 \pm 0.21$	$1.48 \pm 0.03$	$-1.96 \pm 0.06$
0	$0.60 \pm 0.02$	$1.93 \pm 0.06$	$6.14 \pm 0.25$	$1.75 \pm 0.04$	$-1.61 \pm 0.08$
+1	$0.60 \pm 0.02$	$1.99 \pm 0.06$	$6.31 \pm 0.26$	$1.83 \pm 0.04$	$-1.46 \pm 0.09$

Insertion profiles were extended assuming a constant value  $G(z=0)$  in the center of the membrane;  $R^2$ , slope (S), intercept (I), RMSE, and offset values as in Table 2.

**Table 7.**

Evaluation of  $\log P_m$  predictions for corrected  $\log P_m$  PAMPA values with respect to Fujikawa *et al.* for DHDGB and OPLS charges

Profile width change [Å]	$R^2$	Slope	Intercept	RMSE	Offset
Neutral Profile	0.60±0.02	1.62±0.06	5.20±0.26	1.30±0.04	-2.27±0.06
Hybrid Profile ( $G_{state}$ from experimental $pK_a$ )	0.62±0.01	2.17±0.04	6.87±0.18	1.84±0.02	-1.33±0.08
Hybrid Profile ( $G_{state}$ from Leung <i>et al.</i> )	0.66±0.01	2.18±0.05	7.06±0.20	1.74±0.04	-1.49±0.08

$R^2$ , slope (S), intercept (I), RMSE, and offset values as in Table 2.