DEVELOPMENT OF A COMPUTATIONAL APPROACH TO PREDICT BLOOD-BRAIN BARRIER PERMEABILITY

Xingrong Liu, Meihua Tu, Rebecca S. Kelly, Cuiping Chen, AND Bill J. Smith

Department of Pharmacokinetics, Dynamics and Metabolism (X.L., C.C., B.J.S.) and Department of Exploratory Medicinal Sciences (M.T., R.S.K.), Groton Laboratories, Pfizer Global Research and Development, Groton, Connecticut

(Received May 1, 2003; accepted September 2, 2003)

This article is available online at http://dmd.aspetjournals.org

ABSTRACT:

The objectives of this study were to generate a data set of blood-brain barrier (BBB) permeability values for drug-like compounds and to develop a computational model to predict BBB permeability from structure. The BBB permeability, expressed as permeability-surface area product (PS, quantified as logPS), was determined for 28 structurally diverse drug-like compounds using in situ rat brain perfusion technique. A linear model containing three descriptors, logD, van der Waals surface area of basic atoms, and polar surface area, was developed based on 23 compounds in our data set, where the penetration across the BBB was assumed to occur primarily by passive diffusion. The correlation coefficient ($R^2$) and standard deviation (S.D.) of the model-predicted logPS against the observed are 0.74 and 0.50, respectively. If an outlier was removed from the training data set, the $R^2$ and S.D. were 0.80 and 0.44, respectively. This new model was tested in two literature data sets, resulting in an $R^2$ of 0.77 to 0.94 and a S.D. of 0.38 to 0.51. For comparison, four literature models, logP, logD, log(D·MW$^{-0.5}$), and linear free energy relationship, were tested using the set of 23 compounds primarily crossing the BBB by passive diffusion, resulting in an $R^2$ of 0.33 to 0.61 and a S.D. of 0.59 to 0.76. In summary, we have generated the largest PS data set and developed a robust three-descriptor model that can quantitatively predict BBB permeability. This model may be used in a drug discovery setting to predict the BBB permeability of new chemical entities.

The blood-brain barrier (BBB) consists of a continuous layer of endothelial cells joined by tight junctions at the cerebral vasculature. It represents a physical and enzymatic barrier to restrict and regulate the penetration of compounds into and out of the brain and maintain the homeostasis of the brain microenvironment (Davson and Segal, 1995). Brain penetration is essential for compounds where the site of action is within the central nervous system, whereas BBB penetration needs to be minimized for compounds that target peripheral sites to reduce potential central nervous system-related side effects. Therefore, it is critical during the drug discovery phase to select compounds that have appropriate brain penetration properties. Brain penetration is commonly assessed by two experimental approaches, namely equilibrium distribution between brain and blood and BBB permeability. The equilibrium distribution is defined as the ratio of concentrations in brain and blood (BB, quantified as logBB). LogBB is determined as the concentration in brain divided by the concentration in blood. The PS of BBB is used to predict the concentration at steady state or by calculating the area under the brain and blood concentration curve. This parameter is dependent upon the transporters (uptake and efflux) at BBB and the relative drug binding affinity differences between the plasma proteins and brain tissue (Kalvass and Maurer, 2002). A compound with a low logBB value could be attributed to its extensive binding to plasma protein, low partitioning to brain tissue, active efflux at BBB, and/or the sink effect of cerebrospinal fluid. A low logBB value does not necessarily indicate that a compound slowly penetrates across the BBB or is a substrate of an efflux transporter. BBB permeability is often expressed as the BBB permeability-surface area product (PS, quantified as logPS). The PS represents the uptake clearance across the BBB. LogPS can be determined by in vivo intravenous administration, indicator diffusion, brain uptake index techniques, and in situ brain perfusion (Smith, 1989). Unlike logBB, logPS is a direct measure of permeability and theoretically is not confounded by the plasma and brain tissue binding. Therefore, it may be a more relevant parameter to assess the brain penetration properties of a compound in drug discovery.

To design or select compounds with desirable brain penetration properties, there is a great demand in today’s drug discovery setting for rapid and reliable approaches to predict brain penetration for a vast number of compounds. One approach is to use computational methods that can be applied to virtual libraries, allowing rapid and cost-effective elimination of poor candidates, or to rank them even before synthesis, when traditional medicinal chemistry approaches are used. Many computational models using only two-dimensional and three-dimensional structure descriptors have been proposed to predict the logBB and logPS (Levin, 1980; Lombardo et al., 1996; Gratton et al., 1997; Clark, 1999; Lobell et al., 2002; Norinder and Haeberlein, 1997).
Most of the models are developed to predict logBB because logBB is a commonly used parameter to describe brain penetration in drug discovery, and a large logBB database exists in the literature. Since the in vivo determination of BBB permeability is difficult, data on a limited number of compounds have been reported. In addition, because of the variation in in vivo experimental procedures, the various sets of published data may not be comparable and cannot be combined to train and test models. The lack of logPS data has limited the development and validation of models that predict BBB permeability.

In situ brain perfusion developed by Takasato et al. (1984) represents an improved methodology to accurately determine BBB permeability as compared with those in vivo methods discussed above. Several studies have been reported to determine PS values using the brain perfusion method and to examine the correlation of logPS and physical properties. Smith and Takasato (1986) and Gratton et al. (1997) demonstrated that the partitioning coefficient (logP) correlated with BBB permeability. Murakami et al. (2000) observed a correlation of log(D·MW$^{-0.5}$) and logPS for a dozen compounds. Chikhale et al. (1994) suggested that the number of hydrogen bonds was a determinant for BBB permeability for peptide compounds. Using 18 compounds as a training data set, Gratton et al. (1997) proposed a linear free energy relationship (LFER) model. As the first formally proposed logPS model, Gratton and coworkers demonstrated a good correlation between the calculated versus observed logPS values for its training data set with an $R^2$ of 0.97. Nevertheless, this model has not been evaluated using other data sets.

In the present study, the logPS values for 28 drug-like compounds with a range of physicochemical properties were determined. A novel computational model to predict the logPS values was developed, and the performance of the proposed model and four literature models, logP, logD, log(D·MW$^{-0.5}$), and the LFER, in prediction of logPS values was evaluated.

### Materials and Methods

**Reagents.** $[^{14}C]Antipyrine$ (52 mCi/mmol), $[^{14}C]caffeine$ (80 Ci/mmol), $[^{14}C]achloazine$ (55 mCi/mmol), $[^{14}C]dopamine$ (52 mCi/mmol), $[^{14}C]Glycine$ (113 mCi/mmol), $[^{14}C]methotrexate$ (20 Ci/mmol), $[^{14}C]Hephinephrine$ (80 Ci/mmol), $[^{14}C]quinidine$ (20 Ci/mmol), and $[^{14}C]valproic acid$ (55 mCi/mmol) were obtained from American Radiolabeled Chemicals (St. Louis, MO). $[^{14}C]Daunomycin$ (5 Ci/mmol), $[^{14}C]Diazepam$ (83 Ci/mmol), $[^{14}C]Digoxin$ (17 Ci/mmol), $[^{14}C]Phenytoin$ (49 mCi/mmol), $[^{14}C]2.5$-pentenillamine enphalolin (DPDPE) (45 Ci/mmol), $[^{14}C]Phenylalanine$ (51 mCi/mmol), $[^{14}C]Propanolol$ (19 Ci/mmol), $[^{14}C]Cyclic acid$ (56 mCi/mmol), $[^{14}C]Taurocholic acid$ (2 Ci/mmol), $[^{14}C]Testosterone$ (123 Ci/mmol), and $[^{14}C]Xanthine$ (59 mCi/mmol), and were obtained from PerkinElmer Life and Analytical Sciences (Boston, MA). $[^{14}C]SR141716A$ (44 Ci/mmol) was obtained from Amersham Biosciences UK, Ltd. (Little Chalfont, Buckinghamshire, UK). $[^{14}C]Chlorambucil$ (4.9 mCi/mmol), $[^{14}C]Levodopa$ (22 Ci/mmol), $[^{14}C]Hypoxanthine$ (56 mCi/mmol), $[^{14}C]Thiotretinoin$ (56 mCi/mmol), and $[^{14}C]Phenylalanine$ (15 mCi/mmol) were obtained from Moravek Biochemicals (Brea, CA). Fluoxetine was obtained from Sigma-Aldrich (St. Louis, MO). $N^3(3-3^′$-Flurophenyl)-4$^-3^′$-phenyllphenoxy)propylsarcosine (NFPS) and CP-141938 were synthesized at Pfizer Global Research and Development Laboratories (Groton, CT) with purity greater than 98%. Tissue solubilizer Solvable and scintillation cocktail solution Formula-989 were obtained from PerkinElmer Life and Analytical Sciences. All other chemicals used in the experiments were of the highest available grade.

**Brain Perfusion Procedures.** The BBB permeability was determined using a previously reported brain perfusion method (Smith, 1996). In brief, male Sprague-Dawley rats (250–300 g; Charles River Breeding Laboratories, Portage, MI) were anesthetized with ketamine and xylazine (50 mg/kg and 5 mg/kg i.m., respectively). Animal body temperature was maintained at 37°C with a heating pad connected to a temperature controller (TR-100; Fine Science Tools Inc., Foster City, CA). After exposure of the right carotid artery, the right external carotid artery and occipital artery were ligated. Then the right common carotid artery was cannulated with polyethylene tubing (PE-50; Becton Dickinson, Franklin Lakes, NJ) filled with sodium heparin saline (100 IU/ml). After the cannula was in place, the heart was stopped by severing the ventricles, and the infusion was initiated. The animals were perfused for 10 to 60 s at a rate of 10 ml/min using an infusion pump (PHD2000; Harvard Apparatus, Inc., Holliston, MA). The perfusion fluid, maintained at 37°C, was an oxygenated, protein-free, bicarbonate-buffered saline solution containing the following concentrations of solutes: 128 mM NaCl, 4.2 mM KCl, 24 mM NaHCO$_3$, 2.4 mM NaH$_2$PO$_4$, 1.5 mM CaCl$_2$, 0.90 mM MgSO$_4$, and 9.0 mM glucose. $[^{14}C]$ or $[^{3}H]$labeled and nonradiolabeled substrate was added to yield a final perfusion concentration of 0.1 to 1 μM and 0.05 to 5 μM/ml for each test compound. The perfusion was terminated by stopping the pump and decapitating the animal. The excised brain tissue and perfusion fluid were stored at $-20^\circ$C before analysis.

The regional flow rate in the cortex was determined using the PS value of diazepam, a highly permeable compound that has been reported in the literature as a regional flow rate marker in brain perfusion studies (Takasato et al., 1984). Using diazepam, the measured flow rate was 0.070 ± 0.022 ml/g (n = 3) at a 10 ml/min perfusion rate.

**Sample Analysis.** For radiolabeled compounds, frontal and parietal cortex tissues (50–100 mg) were excited, weighed, and placed in a scintillation vial. Samples were digested at 50°C for 12 h in 1 ml of tissue solubilizer. After cooling, the samples were prepared for scintillation counting by addition of 18 ml of scintillation cocktail. To determine the concentration of compound in the perfusion fluid, an aliquot of 20 μl of perfusion fluid was placed in a scintillation vial containing 18 ml of scintillation cocktail. The radioactivity in the brain or perfusion fluid samples was measured with a liquid scintillation counter (Wallac 1409; PerkinElmer Life and Analytical Sciences).

For the analysis of CP-141938, fluoxetine, and NFPS, the brain cortex tissues were homogenized in 3 volumes (w/v) of water. The concentration of CP-141938 in brain homogenate was determined using an HPLC-MS/MS method published previously (Smith et al., 2001). The linear range of quantitation was 2 to 800 ng/g. The fluoxetine and NFPS concentrations in brain homogenate were determined using a modified HPLC-MS/MS method. Briefly, samples of the brain homogenate (0.1 ml) were placed into 1.2-ml Marsh tubes (Marsh Bio Products, Rochester, NY) in a 96-well block and mixed with 0.2 ml of sodium phosphate buffer (pH 7.0). Following the addition of an organic solvent, methyl tert-butyl ether (0.6 ml), the samples were capped and mixed on an orbital shaker for 5 min. The organic and aqueous phases were separated by centrifugation at 3000 rpm for 5 min. The aqueous layer was transferred into a clean microtube using a 19-channel pipette. The organic phase was evaporated to dryness under N$_2$ at 40°C. Sample residues were reconstituted in 50 μl of 60% acetonitrile (v/v) and analyzed by HPLC-MS/MS. The HPLC-MS/MS system consisted of a Shimadzu ternary pump (LC-10A; Shimadzu, Kyoto, Japan), an autosampler, and a PE Sciex API 3000 mass spectrometer with a turbo ion spray interface (PerkinElmerSciex Instruments, Boston, MA). A 10-μl aliquot of each sample was injected onto a Hypurix advanced reverse-phase column (10 × 2 mm, 5 μm; ThermoQuest, Needham, MA). The mobile phase consisted of 10% acetonitrile and 90% 5 mM ammonium formate solution from 0 to 1.0 min. The acetonitrile increased linearly from 10 to 90% from 1.0 to 1.2 min and then was maintained at 90% from 1.2 to 2.0 min. The system returned to the initial conditions in a single linear step and was allowed to equilibrate for 1 min. Fluoxetine and NFPS were eluted at approximately 1.5 min and monitored as the M + H ion conversion at 310.1 → 148.1 and 394.1 → 102.0, respectively. The linear range of quantitation was 8 to 400 ng/g. For all the assays, the precision, expressed as coefficient of variation, was <20%, and the relative accuracy was between 80 and 120%.

**Brain Perfusion Data Analysis.** Calculation of the initial uptake clearance (CL$_{uptake}$/ml/s) for in situ brain perfusion studies was described previously (Smith, 1996; Dagenais et al., 2000). For short-time brain perfusion, the CL$_{uptake}$ was calculated from eq. 1.

$$\text{C}_{\text{Brain}}/\text{C}_{\text{Perfusion}} = \text{CL}_{\text{uptake}} + T + V_s \tag{1}$$

where $C_{\text{Brain}}$ (ng/g) is the brain tissue concentration, $C_{\text{Perfusion}}$ (ng/ml) is the perfusion fluid concentration, $T$ (s) is the perfusion time, and $V_s$ (ml/g) is the
TABLE 1

<table>
<thead>
<tr>
<th>Compound</th>
<th>Mechanism</th>
<th>BB&lt;sub&gt;ko&lt;/sub&gt;/BB&lt;sub&gt;wt&lt;/sub&gt;</th>
<th>Class</th>
<th>Molecular Weight</th>
<th>Calculated Log P</th>
<th>Calculated Log D</th>
<th>(D - MW&lt;sup&gt;0.5&lt;/sup&gt;)&lt;sub&gt;1&lt;/sub&gt;</th>
<th>Vsa_base</th>
<th>TPSA</th>
<th>Observed Log PS</th>
<th>Predicted Log PS</th>
<th>Fold of Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antipyrine</td>
<td>P-gp</td>
<td>Neutral</td>
<td>188.2</td>
<td>0.27</td>
<td>0.27</td>
<td>-0.87</td>
<td>0.00</td>
<td>23.55</td>
<td>-2.0</td>
<td>-2.3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CP-141938</td>
<td>P-gp</td>
<td>50&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Base</td>
<td>403.6</td>
<td>1.13</td>
<td>-1.26</td>
<td>-2.56</td>
<td>11.37</td>
<td>70.67</td>
<td>-3.6</td>
<td>-2.3</td>
<td>18</td>
</tr>
<tr>
<td>SR141716A</td>
<td>P-gp</td>
<td>48.2</td>
<td>Neutral</td>
<td>463.8</td>
<td>4.81</td>
<td>3.48</td>
<td>0.00</td>
<td>50.16</td>
<td>-1.6</td>
<td>-1.5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Caffeine</td>
<td>P-gp</td>
<td>Neutral</td>
<td>194.2</td>
<td>-0.08</td>
<td>-0.08</td>
<td>-1.22</td>
<td>0.00</td>
<td>58.44</td>
<td>-2.0</td>
<td>-2.7</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Chlorambucil</td>
<td>Acid</td>
<td>304.2</td>
<td>3.70</td>
<td>1.13</td>
<td>-0.11</td>
<td>0.00</td>
<td>40.54</td>
<td>-1.6</td>
<td>-2.3</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colchicine</td>
<td>P-gp</td>
<td>3.8&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Neutral</td>
<td>399.4</td>
<td>1.03</td>
<td>1.03</td>
<td>-0.27</td>
<td>0.00</td>
<td>83.09</td>
<td>-3.3</td>
<td>-2.7</td>
<td>4</td>
</tr>
<tr>
<td>DPDPPE</td>
<td>P-gp</td>
<td>2.7&lt;sup&gt;e&lt;/sup&gt;</td>
<td>AA</td>
<td>659.8</td>
<td>0.84</td>
<td>-1.89</td>
<td>-3.30</td>
<td>17.74</td>
<td>199.95</td>
<td>-3.6</td>
<td>-3.2</td>
<td>3</td>
</tr>
<tr>
<td>Daunomycin</td>
<td>P-gp</td>
<td>1.6&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Base</td>
<td>527.5</td>
<td>2.39</td>
<td>0.58</td>
<td>-0.78</td>
<td>17.74</td>
<td>185.84</td>
<td>-2.7</td>
<td>-2.5</td>
<td>1</td>
</tr>
<tr>
<td>Diphenhydramine</td>
<td>P-gp</td>
<td>17&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Neutral</td>
<td>781.0</td>
<td>1.14</td>
<td>1.14</td>
<td>-0.30</td>
<td>0.00</td>
<td>203.06</td>
<td>-4.3</td>
<td>-3.8</td>
<td>3</td>
</tr>
<tr>
<td>Dopamine</td>
<td>P-gp</td>
<td>Neutral</td>
<td>153.2</td>
<td>0.12</td>
<td>-2.09</td>
<td>-3.18</td>
<td>17.74</td>
<td>66.48</td>
<td>-2.1</td>
<td>-2.1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>P-gp</td>
<td>1.1&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Base</td>
<td>309.3</td>
<td>4.35</td>
<td>1.69</td>
<td>0.44</td>
<td>56.8</td>
<td>-1.5</td>
<td>-1.6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Glycine</td>
<td>P-gp</td>
<td>Neutral</td>
<td>138.1</td>
<td>-0.59</td>
<td>-0.59</td>
<td>-1.66</td>
<td>0.00</td>
<td>69.81</td>
<td>-3.5</td>
<td>-3.0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Levodopa</td>
<td>Uptake</td>
<td>AA</td>
<td>197.2</td>
<td>-2.53</td>
<td>-5.04</td>
<td>-6.19</td>
<td>17.74</td>
<td>103.78</td>
<td>-1.8</td>
<td>-3.1</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Methotrexate</td>
<td>Acid</td>
<td>454.5</td>
<td>-0.28</td>
<td>-1.46</td>
<td>-2.79</td>
<td>0.00</td>
<td>210.54</td>
<td>-3.9</td>
<td>-4.5</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphine</td>
<td>P-gp</td>
<td>1.9&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Base</td>
<td>285.3</td>
<td>1.27</td>
<td>0.46</td>
<td>-0.77</td>
<td>0.00</td>
<td>52.93</td>
<td>-2.7</td>
<td>-2.5</td>
<td>2</td>
</tr>
<tr>
<td>NFPS</td>
<td>AA</td>
<td>393.5</td>
<td>5.34</td>
<td>2.84</td>
<td>1.54</td>
<td>0.00</td>
<td>49.77</td>
<td>-2.9</td>
<td>-2.0</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Uptake</td>
<td>AA</td>
<td>165.2</td>
<td>-1.19</td>
<td>-3.70</td>
<td>-4.81</td>
<td>17.74</td>
<td>63.32</td>
<td>-1.3</td>
<td>-2.4</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Phenytoin</td>
<td>P-gp</td>
<td>1.6&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Neutral</td>
<td>252.3</td>
<td>2.52</td>
<td>2.44</td>
<td>1.24</td>
<td>0.00</td>
<td>58.20</td>
<td>-2.2</td>
<td>-2.1</td>
<td>1</td>
</tr>
<tr>
<td>Propranolol</td>
<td>Base</td>
<td>259.4</td>
<td>3.09</td>
<td>1.04</td>
<td>0.13</td>
<td>0.56</td>
<td>41.49</td>
<td>-1.2</td>
<td>-1.8</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quinidine</td>
<td>P-gp</td>
<td>16&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Base</td>
<td>324.4</td>
<td>3.44</td>
<td>1.12</td>
<td>0.06</td>
<td>0.00</td>
<td>45.59</td>
<td>-3.4</td>
<td>-2.2</td>
<td>18</td>
</tr>
<tr>
<td>Salkicylic acid</td>
<td>Acid</td>
<td>138.1</td>
<td>2.06</td>
<td>-1.86</td>
<td>-2.93</td>
<td>0.00</td>
<td>57.53</td>
<td>-3.4</td>
<td>-3.1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taurocholic acid</td>
<td>Acid</td>
<td>515.7</td>
<td>1.04</td>
<td>-3.06</td>
<td>-4.42</td>
<td>0.00</td>
<td>144.16</td>
<td>-4.1</td>
<td>-4.2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testosterone</td>
<td>Neutral</td>
<td>288.4</td>
<td>3.47</td>
<td>3.47</td>
<td>2.24</td>
<td>0.00</td>
<td>37.30</td>
<td>-1.1</td>
<td>-1.7</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theobromine</td>
<td>Base</td>
<td>288.4</td>
<td>-0.79</td>
<td>-0.79</td>
<td>-1.92</td>
<td>0.00</td>
<td>67.23</td>
<td>-3.0</td>
<td>-3.0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theophylline</td>
<td>Base</td>
<td>288.4</td>
<td>0.05</td>
<td>0.02</td>
<td>-1.11</td>
<td>0.00</td>
<td>69.30</td>
<td>-2.9</td>
<td>-2.8</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valproic acid</td>
<td>Acid</td>
<td>144.2</td>
<td>2.93</td>
<td>0.02</td>
<td>-1.10</td>
<td>0.00</td>
<td>37.30</td>
<td>-2.5</td>
<td>-2.5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xanthine</td>
<td>Neutral</td>
<td>152.1</td>
<td>-0.81</td>
<td>-0.81</td>
<td>-1.90</td>
<td>0.00</td>
<td>86.88</td>
<td>-3.8</td>
<td>-3.2</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Brain-to-plasma concentration ratio between the mdr1a or mdr1a/b gene knockout and wild-type mice.
<sup>b</sup> Smith et al. (2001).
<sup>c</sup> Dagenais et al. (2000).
<sup>d</sup> Chen and Pollack (1998).
<sup>e</sup> Adachi et al. (2001).
<sup>f</sup> Schinkel et al. (1995).
<sup>g</sup> Fromm et al. (1999).
<sup>h</sup> Uhr et al. (2000).
<sup>i</sup> Xie et al. (1999).
<sup>j</sup> Ruzic et al. (2002).
<sup>k</sup> Kusuhara et al. (1997).

where F is the regional flow rate estimated from the dicaepm Cl<sub>up</sub> data (0.07 ml/s/g).

**Descriptor Calculations.** Compound structures were sketched into MOE 2002 (http://www.chemcomp.com, Chemical Computing Group, Montreal, Q.B., Canada), and charged and optimized by MMFF94 force field. One hundred six two-dimensional and internal three-dimensional descriptors in MOE 2002 were calculated. LogP and logD (at pH 7.4) were calculated by ACD/LogP 4.56 and ACD/LogD 4.56, respectively (Advanced Chemistry Development, Inc., Toronto, ON, Canada). The logD for DPDPPE and methotrexate was calculated by disabling aliphatic alcohol and amide ionization due to the large number of possible ionization centers in these two molecules. Those most interpretable descriptors that can be directly converted to structures, i.e., fragments and functional groups, were selected using the following approach. First, a correlation coefficient (R<sup>2</sup>) matrix of each pair of the descriptors was generated. Then, all the descriptors with R<sup>2</sup> > 0.80 with molecular weight were removed from the matrix. Similarly, the remaining descriptors with R<sup>2</sup> > 0.80 with logP, TPSA, logD, a_acc, a_don, a_acid, a_base, etc., were removed from the matrix sequentially. Although energy-related descriptors (E<sub>r</sub>) and some partial charge-related descriptors (PEOF_VSA<sup>-r</sup>) are not interpretable descriptors, they were selected due to their lack of correlation with other descriptors. The following 50 descriptors were used in our statistical analysis: logD, logP, a_acc, a_acid, a_aro, a_base, a_don, a_nCl, a_nF, a_nI, a_nN, b_double, b_rotN, b_rotR, density, dipole, E_ang, E_ele, E_nb, E_ogp, E_sol, E_sub, E_str, E_tor, PEOR_RPC+, PEOR_E-RPC+, PEOR_VSA+, PEOR_VSA-1, PEOR_VSA-2, PEOR_VSA-3, PEOR_VSA+, PEOR_VSA+, PEOR_VSA+, PEOR_VSA+, PEOR_VSA+, PEOR_VSA-, PEOR_VSA-, PEOR_VSA-, PEOR_VSA-, PEOR_VSA+, PEOR_VSA+, PEOR_VSA+. Model and regression equations were obtained by multivariate regression analysis in JMP 4.0 (SAS Institute, Inc., Cary, NC). The fold of error, which is an index for the prediction of accuracy obtained by the ratio of the predicted and observed estimates, was calculated for each compound as follows: fold of error = PS<sub>predicted</sub>/PS<sub>observed</sub> if PS<sub>predicted</sub> > PS<sub>observed</sub>; otherwise, fold of error = PS<sub>observed</sub>/PS<sub>predicted</sub>.

**Results.** Log PS and Descriptors. The determined logPS values for 28 compounds and the available literature on BB<sub>ko</sub>/BB<sub>wt</sub> values are presented in Table 1. BB<sub>ko</sub>/BB<sub>wt</sub> represents the ratio of the BB obtained from mdr1a or mdr1a/b gene-knockout mice and the BB.
compounds. The open squares and open triangles represent the active uptake and efflux substrates (BBKO/BB WT ratio

$H_11002^{2.0}$ and S.D. will be 0.80 and 0.44, respectively. For uptake substrates (alanine, phenylalanine, and glucose), efflux substrates (vincristine, vinblastine, quinidine, cyclosporine A, and digoxin) as suggested by Murakami et al. (2000), and an outlier, cimetidine, from the original 21 compounds (eq. 5).

For Murakami’s data set:

$$\logPS_{\text{Predicted}} = -0.161 + 0.950 \logPS_{\text{Observed}}$$

(5)

In Gratton’s data set, six ionized compounds at pH 7.4 were excluded from this testing data set. A good correlation was observed for the remaining 12 neutral compounds (eq. 4). For Murakami’s data set, the predicted logPS values were well correlated with the observed values after excluding uptake substrates (alanine, phenylalanine, and levodopa, whose logPS values were significantly overpredicted by 3-, 18-, and 18-fold, respectively.

**Model Evaluation.** Our proposed model was tested against two literature data sets. The first data set contains 12 compounds reported by Gratton et al. (1997), and the second data set includes 13 compounds reported by Murakami et al. (2000). The observed values and the predicted values by using eq. 3 and the fold of error are listed in Tables 2 and 3 and are presented in Fig. 2.

For Gratton’s data set:

$$\logPS_{\text{Predicted}} = 1.17 + 0.639 \logPS_{\text{Observed}}$$

(4)

$$R^2 = 0.94, \text{S.D.} = 0.38, F = 159, n = 12.$$
sent in Fig. 3D. A simple linear regression of the logPS of the 10 neutral diffusion compounds against predicted logPS yields eq. 10:

$$\log_{\text{PS,Observed}} = -1.67 + 0.463 \log_{\text{PS,Predicted}} \quad (10)$$

$$R^2 = 0.61, \text{ S.D.} = 0.59, F = 12.4, n = 10.$$  

**Discussion**

The objectives of this study were to generate a data set of BBB permeability values for drug-like compounds and to develop a computational model to predict BBB permeability from structure. The development of in silico logPS models has been hindered by the lack of a large drug-like logPS data set. So far, only a few small logPS data sets have been published. In the published data sets, many non-drug-like compounds including solvents, such as ethanol and propanol, and sugars, such as sucrose and mannitol, were included (Smith and Takasato, 1986; Gratton et al., 1997; Murakami et al., 2000). In the present study, we determined logPS values for 28 drug-like compounds using the in situ brain perfusion method. Our logPS data set is unique in the range of structurally diverse drug-like compounds. No solvents have been included in our data set. Therefore, this data set more closely resembles those compounds that are encountered in today’s drug discovery programs. Addition of our logPS data hopefully will stimulate the research to develop better logPS models.

Two observations were made from the logPS data. First, disregarding the physical properties of compounds, the order of BBB permeability is active uptake compounds > passive diffusion compounds > efflux compounds. The average logPS of the two active uptake substrates is approximately one log unit greater than that of passive diffusion substrates; the average of logPS of diffusion substrates is approximately one log unit greater than that of efflux (P-gp) substrates. Second, basic compounds appear to have higher BBB permeability than neutral and acidic compounds in this data set. The rank order of the average logPS values for the passive diffusion compounds by ionization charge status is basic compounds > neutral compounds > acidic compounds.

Initial attempts to build a general model for all 28 compounds were unsuccessful. Due to the complexity of substrate structure and transport-activity relationships, a general computational model, which relies on relative generic property descriptors, will not likely be able to predict brain uptake for both simple diffusion compounds and actively transported compounds. We refined the logPS model to predict the BBB permeability for 23 substrates that penetrate BBB primarily through a passive diffusion mechanism. Phenylalanine and levodopa have been reported as good substrates of system L, for large neutral amino acids, an uptake transport system at BBB (Wade and Katzman, 1975; Momma et al., 1987; Alexander et al., 1994; Sanchez del Pino et al., 1995). According to $BB_{Ko} / BB_{WT}$ ratio > 2, five compounds in our data set have been reported as possible substrates for P-gp, an
efflux transport system at BBB (Table 1). Among the five compounds, CP-141938, digoxin, and quinidine show BB KO /BB WT ratio greater than 10 and are classified as good P-gp substrates. The other two substrates (colchicine and DPDPE) show BBKO /BB WT ratio between 2.7 and 3.8. If we assume that the BB KO /BB WT ratio correlates with the effect of P-gp on logPS value, the BBB permeability of these two compounds may be 2- to 4-fold lower than that of those diffusion compounds that have similar physicochemical properties. Considering the variability of the BBKO /BB WT ratios, our experiment for PS measurement, and the accuracy of a computational approach, it is unlikely that our model will be able to differentiate 2- to 4-fold differences for the BBB permeability. In addition, in the plots of logPS versus logP, logD, log(D /MW ^0.5), and the plot of calculated logPS from LFER model versus observed logPS, colchicine and DPDPE do not show an obvious separation from the passive diffusion compounds. These two compounds were treated as passive diffusion substrates for the purpose of model training and testing. Therefore, two uptake and three efflux substrates have been excluded from the model training, and the remaining 23 compounds were classified as passive diffusion substrates and used for testing literature models and training our model.

In the present study, among 106 descriptors that have been examined, three common structure descriptors, logD, TPSA, and vsa_base, were identified as the most important descriptors that correlate with BBB permeability. A linear model containing these three descriptors has been developed (eq. 3). These descriptors can be readily calculated using commercially available software. Therefore, this model is suitable in a drug discovery setting to predict logPS values in a moderate- to high-throughput mode. The bottleneck of the calculation is the estimation of logD. The speed of the calculation is approximately 100 compounds/min. Consistent with literature reports, logD, a measure of lipophilicity, has positive contribution to BBB permeability. TPSA measures a compound’s polarity and hydrogen bonding potential and has negative contribution to logPS. A larger TPSA value usually deters a compound from entering the brain. This observation agrees with the work reported by Chikhale et al. (1994), who have demonstrated that the number of hydrogen bonds is a determinant for BBB permeability for seven peptide compounds. Descriptor vsa_base represents van der Waals surface area of basic atoms and indicates the basicity of a compound. Under physiological conditions, a compound with a high vsa_base value tends to be protonated and carries positive charges. The positive contribution to logPS indicates that basicity facilitates the compound’s permeability across the BBB. This is consistent with our observation that basic compounds have higher logPS values than neutral and acidic compounds.

Our model can describe the 23 diffusion compounds in the training...
data sets and can differentiate uptake and efflux substrates. The predicted logPS values correlated with the observed values ($R^2 = 0.74$, S.D. = 0.50) (Fig. 1). The percentage of compounds within 2- and 4-fold of error is 52% (12 of 23) and 83% (19 of 23), respectively. NFPs had the largest discrepancy between predicted and observed PS in our model (fold of error = 9). If this compound is excluded from the training data, the $R^2$ and S.D. of our model will be 0.80 and 0.44, respectively. However, due to lack of direct evidence to demonstrate that NFPs is indeed a substrate for a known efflux transporter, NFPs was included in the training data set as a passive diffusion compound to remove bias from our model-building process. Our model appears to be able to differentiate good uptake and efflux substrates from those compounds whose brain penetration is primarily governed by a passive diffusion mechanism. The four compounds with the highest fold of error (12- to 18-fold) are either good uptake substrates (levodopa, theophylline, and phenylalanine) or good efflux transporter P-gp substrates (CP-141938 and quinidine) (Fig. 1).

Our proposed model can rank the logPS for the compounds reported by Gratton et al. (1997). The reported PS values by Gratton et al. (1997) were corrected for the neutral species by considering the fraction ionized at physiological pH calculated from their respective $pK_a$ values (M. Abraham, personal communication, 2003). Since the $pK_a$ values for six ionized compounds are not available, these six compounds were excluded. For the remaining 12 compounds, a good correlation is obtained between the predicted and reported logPS ($R^2 = 0.94$, S.D. = 0.38). Although our model can rank the compounds, it has a large error for quantitative prediction. The percentage of compounds within 2- and 4-fold of error is 25% (3 of 12) and 67% (8 of 12), respectively. Many of the compounds with large fold of error are simple organic solvents, such as ethanol and 2-propanol, and their physicochemical properties are quite different from the compounds in our model training data set. The prediction errors may be also caused by a systematic difference between Gratton’s data and our data, since the intercept (1.17) and slope (0.639) of the regression line of the predicted versus observed data are significantly different from zero and unity, respectively (eq. 4).

The proposed model can quantitatively predict the log PS for the compounds reported by Murakami et al. (2000). A good correlation of predicted and reported logPS values was observed for 12 passive diffusion compounds that were suggested by the authors (Fig. 2B) ($R^2 = 0.77$, S.D. = 0.51). The percentage of compounds within 2- and 4-fold of error is 42% (5 of 12) and 92% (11 of 12), respectively. The lack of a systematic difference between these two data sets contributes to the success of prediction. The intercept ($-0.161$) for Murakami’s data set is close to zero and the slope (0.950) is close to one. In addition, sharing of a subset of data between these two data sets may also contribute to the success of prediction. For the seven compounds that exist in both data sets, an excellent correlation was observed (logPS$_{Murakami}$’s data = $0.158 + 1.04$ logPS$_{Liu}$’s data, $R^2 = 0.91$, S.D. = 0.31). We noticed that our model underpredicts a P-gp substrate, cyclosporine A, and overpredicts an organic anion inhibitor, cimetidine, by 120-fold. The discrepancy for cyclosporine A could be due to either experimental error in the observed logPS values or lack of similar compounds in our training data set. The overprediction of the cimetidine PS value could be due to decrease in BBB permeability by interacting with endothelial H$_2$ receptors (Butt and Jones, 1992) or an unidentified efflux transporter at BBB, since cimetidine has been reported as a substrate of an organic anion transporter (Oat 3) at the choroid plexus (Kusuhara et al., 1999; Nagata et al., 2002).

Simple descriptors such as logP, logD, and logD (MW$^{-0.5}$) correlate with logPS values for our data set to some extent; nevertheless, none of these correlations are sufficient to quantitatively predict the logPS values ($R^2 = 0.33–0.45$, S.D. = 0.69–0.76). Although a good correlation of logP versus logPS was reported by Smith and Takasato (1986) ($R^2 = 0.96$, $n = 12$) and Gratton et al. (1997) ($R^2 = 0.88$, $n = 18$), log P did not correlate with the log PS in our data set ($R^2 = 0.33$, $n = 23$). However, an improved correlation was observed for the neutral compounds in our data set ($R^2 = 0.55$, $n = 10$). Considering the wide range of $pK_a$ values of our model compounds, logD might be a better descriptor to describe lipophilicity under physiological conditions as compared with logP, since logP only counters for the neutral form and log D counters for both neutral and ionized forms. A better correlation ($R^2 = 0.44$) between log D and log PS is indeed observed. Murakami et al. (2000) have reported that logD (MW$^{-0.5}$) appears to correlate with logPS values for a set of diffusion compounds, although no statistical analysis is provided in their study. Clearly, in our data set molecular weight does not improve the correlation of logD with logPS. Using logD (MW$^{-0.5}$) instead of logD, the $R^2$ of the correlation only slightly increased from 0.44 to 0.45.

The LFER model proposed by Gratton et al. (1997) cannot quantitatively predict our logPS data set. The LFER model was intended to predict the permeability of the neutral species; therefore, only the 11 neutral compounds in our data set were selected to test the LFER model. Although an excellent correlation between the calculated and observed logPS values was reported for Gratton’s data set ($R^2 = 0.98$, $n = 18$), poor correlation was observed for the neutral compounds in our data set ($R^2 = 0.60$, $n = 10$). The percentage of compounds within 2- and 4-fold of error is 18% (2 of 11) and 55% (6 of 11), respectively. Several factors may cause the failure for Gratton’s model to predict our data set. First, only a limited number of compounds were used to train the model, and the chemical space covered by the original model might not be a good representation of our data set. Second, as discussed earlier, a significant systematic difference in PS values exists between these two data sets. Third, it is difficult to estimate Abraham’s descriptors due to intramolecular interactions between many

### Table 4

<table>
<thead>
<tr>
<th>Compound</th>
<th>$\Sigma^O$</th>
<th>$\Sigma^T$</th>
<th>$R_s$</th>
<th>$V_s$</th>
<th>Observed Log PS</th>
<th>Predicted Log PS</th>
<th>Fold of Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antipyrine</td>
<td>1.33</td>
<td>1.55</td>
<td>1.45</td>
<td>1.48</td>
<td>-2.0</td>
<td>-1.8</td>
<td>1</td>
</tr>
<tr>
<td>SR141716A</td>
<td>1.92</td>
<td>3.63</td>
<td>3.11</td>
<td>3.21</td>
<td>-1.6</td>
<td>-0.36</td>
<td>17</td>
</tr>
<tr>
<td>Caffeine</td>
<td>1.48</td>
<td>1.44</td>
<td>1.49</td>
<td>1.36</td>
<td>-2.0</td>
<td>-2.4</td>
<td>3</td>
</tr>
<tr>
<td>Colchicine</td>
<td>2.02</td>
<td>3.26</td>
<td>2.15</td>
<td>2.99</td>
<td>-3.3</td>
<td>-1.4</td>
<td>79</td>
</tr>
<tr>
<td>Diclofenac</td>
<td>4.75</td>
<td>5.46</td>
<td>3.25</td>
<td>5.75</td>
<td>-4.3</td>
<td>-3.2</td>
<td>3</td>
</tr>
<tr>
<td>Hypoxanthine</td>
<td>1.41</td>
<td>1.94</td>
<td>1.51</td>
<td>0.925</td>
<td>-3.5</td>
<td>-4.6</td>
<td>13</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>1.14</td>
<td>1.86</td>
<td>2.14</td>
<td>1.87</td>
<td>-2.2</td>
<td>-0.050</td>
<td>141</td>
</tr>
<tr>
<td>Testosterone</td>
<td>1.19</td>
<td>2.35</td>
<td>1.55</td>
<td>2.38</td>
<td>-1.1</td>
<td>0.13</td>
<td>17</td>
</tr>
<tr>
<td>Theobromine</td>
<td>1.35</td>
<td>1.60</td>
<td>1.60</td>
<td>1.22</td>
<td>-3.0</td>
<td>-2.7</td>
<td>3</td>
</tr>
<tr>
<td>Theophylline</td>
<td>1.42</td>
<td>1.36</td>
<td>1.49</td>
<td>1.22</td>
<td>-2.9</td>
<td>-2.5</td>
<td>3</td>
</tr>
<tr>
<td>Xanthine</td>
<td>1.24</td>
<td>1.56</td>
<td>1.58</td>
<td>0.941</td>
<td>-3.8</td>
<td>-3.3</td>
<td>3</td>
</tr>
</tbody>
</table>
functional groups (Norinder and Haeberlein, 2002). Nevertheless, with the addition of our data set, a more accurate LFER model may be developed.

The proposed three-descriptor model clearly demonstrates its superiority in logPS prediction as compared with several previously reported literature models. This model can be further refined when PS data for additional compounds are available. Our work indicates that there are two utilities for the proposed model in drug discovery. It helps the design of compounds with high, passive BBB permeability by predicting or ranking compounds according to their BBB permeability at early stages of drug discovery. The other utility is to assess whether an observed logPS can be explained by its physicochemical properties. If the predicted logPS is significantly greater or lower than the observed value, this indicates the compound may be a good efflux substrate or a good uptake substrate, respectively. This approach is particularly valuable to address the situation where genetically modified mice for newly identified transporters are not available.

In summary, a logPS data set of 28 drug-like compounds obtained under consistent experimental conditions has been reported, and a three-descriptor model proposed in this report can predict or rank logPS for compounds not existing in the training data set. This model may be used in a drug discovery setting to predict the BBB permeability of new chemical entities.

Acknowledgments. We thank Dr. Quentin Smith for providing training on the perfusion method, helpful discussions, and valuable comments. We also thank Dr. Michael Abraham for the calculation of Abraham’s descriptors, and JianHua Liu and Angela Doran for the HPLC-MS/MS sample analyses.

References


