Supporting Information:

A computational scheme of pK\textsubscript{a} values based on the three-dimensional reference interaction site model self-consistent field theory coupled with the linear fitting correction scheme

*Ryo Fujiki\textsuperscript{1}, Yukako Kasai\textsuperscript{1}, Yuki Seno\textsuperscript{1}, Toru Matsui\textsuperscript{2}, Yasuteru Shigeta\textsuperscript{3}, Norio Yoshida*\textsuperscript{*1}, Haruyuki Nakano\textsuperscript{1}*

1. Department of Chemistry, Graduate School of Science, Kyushu University, 744, Motooka, Nishiku, Fukuoka, 819-0395 Japan

2. Department of Chemistry, Graduate School of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577 Japan

3. Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577 Japan
Table S1(1). Training set molecules for the LFC parameter fitting and those experimental $pK_a$ values.

<table>
<thead>
<tr>
<th>Carboxyl</th>
<th>Molecule</th>
<th>$pK_a$</th>
<th>Ref.</th>
<th>Amine</th>
<th>Molecule</th>
<th>$pK_a$</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHOOCOOH</td>
<td>3.32</td>
<td>1</td>
<td></td>
<td>Ph(CH$_2$)$_3$NH$_3^+$</td>
<td>9.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>trans-CH$_3$CH=CHCOOH</td>
<td>4.69</td>
<td></td>
<td></td>
<td>PhCH$_2$NH$_3^+$</td>
<td>9.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ph(OH)$_2$COOH</td>
<td>4.48</td>
<td></td>
<td></td>
<td>PhNH$_3^+$</td>
<td>4.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H$_2$C=CHCH$_2$COOH</td>
<td>4.42</td>
<td></td>
<td></td>
<td>CH$_3$(CH$_2$)$_3$NH$_3^+$</td>
<td>10.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(CH(OH)COOH)$_2$</td>
<td>1.14</td>
<td>2</td>
<td></td>
<td>CH$_3$CH$_2$NH$_3^+$</td>
<td>10.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHOCH$_3$COOH</td>
<td>3.86</td>
<td></td>
<td></td>
<td>HO(CH$_2$)$_3$NH$_3^+$</td>
<td>9.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH$_3$COCH$_2$COOH</td>
<td>3.58</td>
<td></td>
<td></td>
<td>HONH$_3^+$</td>
<td>5.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH$_3$COCOOH</td>
<td>2.50</td>
<td>3</td>
<td></td>
<td>NH$_4^+$</td>
<td>9.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHCl$_2$COOH</td>
<td>1.29</td>
<td></td>
<td></td>
<td>H$_2$C=CHCH$_2$NH$_3^+$</td>
<td>9.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH$_2$FCOOH</td>
<td>2.66</td>
<td></td>
<td></td>
<td>CH$_3$(CH$_2$)$_2$NH$_3^+$</td>
<td>10.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO$_2$CH$_2$COOH</td>
<td>1.68</td>
<td></td>
<td></td>
<td>Cyclohexylamine</td>
<td>10.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PhNO$_2$COOH</td>
<td>2.45</td>
<td>4</td>
<td></td>
<td>Cyclohexylmethyl amine</td>
<td>10.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Isopropylamine</td>
<td>10.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Methoxyamine</td>
<td>4.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\gamma$-Phenylpropyl amine</td>
<td>10.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>neo-Pentylamine</td>
<td>10.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sec-Butylamine</td>
<td>10.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S1(2). Training set molecules for the LFC parameter fitting and those experimental pKa values.

<table>
<thead>
<tr>
<th>Imidazole</th>
<th>Molecule</th>
<th>pK\textsubscript{a}</th>
<th>Ref.</th>
<th>Molecule</th>
<th>pK\textsubscript{a}</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imidazole</td>
<td>2-Methyl-4-hydroxy-aminobenzimidazole</td>
<td>6.65</td>
<td></td>
<td>C\textsubscript{2}H\textsubscript{5}OCH\textsubscript{2}CH\textsubscript{2}SH</td>
<td>9.38</td>
<td>6</td>
</tr>
<tr>
<td>Imidazole</td>
<td>2-Methylbenzimidazole</td>
<td>6.10</td>
<td></td>
<td>C\textsubscript{2}H\textsubscript{5}OCOCH\textsubscript{2}SH</td>
<td>7.95</td>
<td>7</td>
</tr>
<tr>
<td>Imidazole</td>
<td>2-Methylimidazole</td>
<td>7.75</td>
<td></td>
<td>C\textsubscript{2}H\textsubscript{5}SH</td>
<td>9.43</td>
<td></td>
</tr>
<tr>
<td>Imidazole</td>
<td>4-Hydroxy-6-aminobenzimidazole</td>
<td>5.90</td>
<td></td>
<td>CH\textsubscript{3}=CHCH\textsubscript{2}SH</td>
<td>9.96</td>
<td></td>
</tr>
<tr>
<td>Imidazole</td>
<td>4-Hydroxy benzimidazole</td>
<td>5.30</td>
<td></td>
<td>HOC\textsubscript{2}CHOHCH\textsubscript{2}SH</td>
<td>9.51</td>
<td></td>
</tr>
<tr>
<td>Imidazole</td>
<td>4-Methoxy benzimidazole</td>
<td>5.10</td>
<td></td>
<td>n-C\textsubscript{2}H\textsubscript{5}SH</td>
<td>10.65</td>
<td></td>
</tr>
<tr>
<td>Imidazole</td>
<td>4-Nitroimidazole</td>
<td>1.50</td>
<td></td>
<td>n-C\textsubscript{4}H\textsubscript{9}SH</td>
<td>10.66</td>
<td>8</td>
</tr>
<tr>
<td>Imidazole</td>
<td>6-Nitrobenzimidazole</td>
<td>3.05</td>
<td></td>
<td>t-C\textsubscript{4}H\textsubscript{9}SH</td>
<td>11.21</td>
<td></td>
</tr>
<tr>
<td>Imidazole</td>
<td>Benzimidazole</td>
<td>5.40</td>
<td></td>
<td>2-Mercaptoethanol</td>
<td>9.50</td>
<td></td>
</tr>
<tr>
<td>Imidazole</td>
<td>2-Methyl-4-hydroxy-6-nitrobenzimidazole</td>
<td>3.90</td>
<td></td>
<td>2-Mercaptoethylamine</td>
<td>8.60</td>
<td></td>
</tr>
<tr>
<td>Imidazole</td>
<td>4-Hydroxy-6-nitrobenzimidazole</td>
<td>3.05</td>
<td></td>
<td>Thioglycolic acid</td>
<td>10.31</td>
<td>9</td>
</tr>
<tr>
<td>Imidazole</td>
<td>4-(2-4-dihydroxy phenyl)-imidazole</td>
<td>6.45</td>
<td></td>
<td>Thiophenol</td>
<td>7.8</td>
<td></td>
</tr>
<tr>
<td>Imidazole</td>
<td>4-Methyl-imidazole</td>
<td>7.45</td>
<td></td>
<td>o-Aminothiophenol</td>
<td>6.59</td>
<td></td>
</tr>
<tr>
<td>Imidazole</td>
<td>6-Aminobenzidazole</td>
<td>6.00</td>
<td></td>
<td>3-Mercaptopropionic acid</td>
<td>10.27</td>
<td></td>
</tr>
<tr>
<td>Imidazole</td>
<td>Histamine</td>
<td>6.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S1(3). Training set molecules for the LFC parameter fitting and those experimental $pK_a$ values.

<table>
<thead>
<tr>
<th>Alcohol</th>
<th>Molecule</th>
<th>$pK_a$</th>
<th>Ref.</th>
<th>Phenol</th>
<th>Molecule</th>
<th>$pK_a$</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CCl$\text{3}$CH$\text{2}$OH</td>
<td>11.80</td>
<td>10</td>
<td></td>
<td>2Cl-4NO$\text{2}$-phenol</td>
<td>5.42</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>CHF$\text{3}$CF$\text{2}$CH$\text{2}$OH</td>
<td>11.34</td>
<td></td>
<td></td>
<td>C$\text{3}$H$\text{2}$CH$\text{2}$O$\text{2}$-phenol</td>
<td>8.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CH$\text{2}$=CHCH$\text{2}$OH</td>
<td>15.10</td>
<td></td>
<td></td>
<td>m-CH$\text{3}$CO-phenol</td>
<td>9.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CH$\text{2}$CH$\text{2}$OH</td>
<td>15.90</td>
<td></td>
<td></td>
<td>m-CH$\text{2}$O-phenol</td>
<td>9.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CH$\text{3}$OCH$\text{2}$CH$\text{2}$OH</td>
<td>14.80</td>
<td></td>
<td></td>
<td>m-F-phenol</td>
<td>9.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CH$\text{3}$OH</td>
<td>15.54</td>
<td></td>
<td></td>
<td>m-HOCH$\text{2}$-phenol</td>
<td>9.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CHCCH$\text{2}$OH</td>
<td>13.55</td>
<td></td>
<td></td>
<td>m-NH$\text{2}$-phenol</td>
<td>9.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CHCl$\text{3}$CH$\text{2}$OH</td>
<td>12.89</td>
<td></td>
<td></td>
<td>o-OCH-phenol</td>
<td>6.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HOCH$\text{2}$CF$\text{2}$CH$\text{2}$OH</td>
<td>11.00</td>
<td></td>
<td></td>
<td>p-Br-phenol</td>
<td>9.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CH$\text{3}$OCH$\text{2}$OH</td>
<td>14.80</td>
<td>11</td>
<td></td>
<td>p-C$\text{3}$H$\text{5}$O$\text{2}$-phenol</td>
<td>8.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C(CH$\text{2}$OH)$\text{4}$</td>
<td>14.10</td>
<td></td>
<td></td>
<td>p-C$\text{6}$H$\text{5}$-phenol</td>
<td>9.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HOCH$\text{2}$CHOHCH$\text{2}$OH</td>
<td>14.40</td>
<td></td>
<td></td>
<td>p-CH$\text{3}$O$\text{2}$-phenol</td>
<td>8.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C$\text{3}$H$\text{2}$OH</td>
<td>16.00</td>
<td></td>
<td></td>
<td>p-CH$\text{3}$S-phenol</td>
<td>9.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CF$\text{3}$CH$\text{2}$OH</td>
<td>12.37</td>
<td></td>
<td></td>
<td>p-CH$\text{3}$SO$\text{2}$-phenol</td>
<td>7.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HOCH$\text{2}$CH$\text{2}$OH</td>
<td>14.77</td>
<td></td>
<td></td>
<td>p-HO-phenol</td>
<td>9.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CF$\text{3}$C(CH$\text{3}$)$\text{2}$OH</td>
<td>11.60</td>
<td></td>
<td></td>
<td>p-NC-phenol</td>
<td>7.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CF$\text{3}$CH(OH)CH$\text{3}$</td>
<td>11.80</td>
<td>12</td>
<td></td>
<td>p-O$\text{2}$C-phenol</td>
<td>9.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p-(CH$\text{3}$)N$^+$phenol</td>
<td>8.00</td>
<td></td>
</tr>
</tbody>
</table>
Table S2. The fitted parameters, RMSE, correlation factor \( r \) and \( G(H^+) \) for LFC/PCM in each chemical group, using 6-31++G(d,p).

<table>
<thead>
<tr>
<th></th>
<th>( k^a )</th>
<th>( C_0 )</th>
<th>( s )</th>
<th>RMSE</th>
<th>( r )</th>
<th>( G(H^+)^b )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol</td>
<td>0.238</td>
<td>-61.002</td>
<td>0.325</td>
<td>1.115</td>
<td>0.733</td>
<td>-255.9</td>
</tr>
<tr>
<td>Amine</td>
<td>0.380</td>
<td>-98.353</td>
<td>0.519</td>
<td>0.597</td>
<td>0.955</td>
<td>-258.8</td>
</tr>
<tr>
<td>Imidazole</td>
<td>0.310</td>
<td>-81.673</td>
<td>0.423</td>
<td>0.647</td>
<td>0.922</td>
<td>-263.2</td>
</tr>
<tr>
<td>Thiol</td>
<td>0.260</td>
<td>-66.928</td>
<td>0.354</td>
<td>0.866</td>
<td>0.716</td>
<td>-257.8</td>
</tr>
<tr>
<td>Phenol</td>
<td>0.223</td>
<td>-57.650</td>
<td>0.305</td>
<td>0.441</td>
<td>0.925</td>
<td>-258.0</td>
</tr>
<tr>
<td>Carboxyl</td>
<td>0.221</td>
<td>-60.102</td>
<td>0.302</td>
<td>0.622</td>
<td>0.853</td>
<td>-271.4</td>
</tr>
<tr>
<td>Total</td>
<td>0.721</td>
<td>0.977</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Unit of \( k \) is mol/kcal. \(^b\) Unit of \( G(H^+) \) is kcal/mol.
Table S3. The comparison of the computed and experimental $pK_a$ values for amino acids using the PCM.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Chemical group</th>
<th>$pK_a$</th>
<th>LFC/PCM</th>
<th>direct PCM</th>
<th>expt$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp</td>
<td>Carboxyl</td>
<td>1.97</td>
<td>27.51</td>
<td>3.86</td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>Thiol</td>
<td>7.10</td>
<td>31.10</td>
<td>8.33</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Carboxyl</td>
<td>4.25</td>
<td>35.08</td>
<td>4.25</td>
<td></td>
</tr>
<tr>
<td>His(D/E)$^b$</td>
<td>Imidazole</td>
<td>6.23/6.23</td>
<td>29.72/29.74</td>
<td>6.04</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Amine</td>
<td>12.02</td>
<td>34.96</td>
<td>10.53</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Phenol</td>
<td>9.96</td>
<td>43.85</td>
<td>10.07</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Taken from ref. 15. $^b$ D and E denote the positions where protonation occurs, the delta and epsilon nitrogens, respectively.
References


