

# Improving the Accuracy of AutoDock Vina by Changing the Empirical Parameters

T. Ngoc Han Pham,<sup>a</sup> Trung Hai Nguyen,<sup>bc</sup> Nguyen Minh Tam,<sup>d</sup> Thien Y Vu,<sup>a</sup> Nhat Truong Pham,<sup>e</sup> Nguyen Truong Huy,<sup>a</sup> Binh Khanh Mai,<sup>f</sup> Nguyen Thanh Tung,<sup>gh</sup> Minh Quan Pham,<sup>hi</sup> Van V. Vu,<sup>j</sup> and Son Tung Ngo<sup>bc\*</sup>

<sup>a</sup>Faculty of Pharmacy, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam

<sup>b</sup>Laboratory of Theoretical and Computational Biophysics, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam

<sup>c</sup>Faculty of Applied Sciences, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam

<sup>d</sup>Computational Chemistry Research Group, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam

<sup>e</sup>Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City 72915, Vietnam

<sup>f</sup>Department of Chemistry, University of Pittsburgh, Pittsburgh, PA, 15260, USA

<sup>g</sup>Institute of Materials Science, Vietnam Academy of Science and Technology, Hanoi 11355, Vietnam

<sup>h</sup>Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi 11355, Vietnam

<sup>i</sup>Institute of Natural Products Chemistry, Vietnam Academy of Science and Technology, Hanoi 11355, Vietnam

<sup>j</sup>NTT Hi-Tech Institute, Nguyen Tat Thanh University, Ho Chi Minh City 72820, Viet Nam

**ABSTRACT:** According to the previous benchmark, Autodock Vina (Vina) achieved a very high successful-docking rate,  $\hat{p}$ , but give a rather a low correlation coefficient,  $R$ , for binding affinity with respect to experiment. This low correlation can be an obstacle for ranking of ligand-binding affinity, which is a main objective of docking simulations. The accuracy of Vina likely depends on the empirical parameters, which include the Gaussian steric interaction, repulsion, hydrophobic, hydrogen bond, and rotation metrics. In this context, we evaluated the dependence of Vina accuracy upon empirical parameters. Although changing of six parameters alters the obtained  $R$  values, the *gauss2* and *rotation* terms form more effects. The  $\hat{p}$  terms are sensitive to the alterations of the *gauss1*, *gauss2*, *repulsion*, and *hydrogen bond* parameters. Therefore, three sets of empirical parameters were proposed as well as more to modestly focus on  $R$  (set1), modestly focus on  $\hat{p}$  (set3), and keep a tradeoff between  $R$  and  $\hat{p}$ . The testing study over 800 complexes indicated that the Vina with proposed sets of parameters can provide more accurate results since forming the larger  $R$  value ( $R_{\text{set1}} = 0.556 \pm 0.025$ ) compared with the original one ( $R_{\text{Default}} = 0.493 \pm 0.028$ ) and Vina version 1.2 ( $R_{\text{Vina 1.2}} = 0.503 \pm 0.029$ ). Besides, the testing study over 48 biological targets indicated that the modified Vina can be applied more widely compared with the default package. These newly proposed parameters achieved a higher correlation coefficient and reasonable correlation coefficients ( $R > 0.500$ ) for at least 32 targets, whereas the default parameters provided by the original Vina gave only 31 targets with at least 0.500 correlation. In addition, validation calculations for 1315 complexes obtained from the version 2019 of PDBbind refined

structures suggested that set1 of parameters are more appropriate than the other parameters ( $R_{\text{set1}} = 0.621 \pm 0.016$ ) compared with the default package ( $R_{\text{Default}} = 0.552 \pm 0.018$ ) and Vina version 1.2 ( $R_{\text{Vina 1.2}} = 0.549 \pm 0.017$ ). The version of Vina with set1 of parameters can be downloaded at <https://github.com/sontungngo/mvina>. The outcomes probably enhance the ranking of ligand-binding affinity using Autodock Vina.

## INTRODUCTION

The ligand-binding process is one of the most important issues in biology.<sup>1</sup> These processes are mostly associated with noncovalent chemical reactions between inhibitors and protein targets.<sup>2</sup> In particular, the process can be mimicked using computational approaches,<sup>3,4</sup> which plays a tremendous role in computer-aided drug design (CADD).<sup>5</sup> Accurate determination of ligand-binding affinity and pose of a small compound to an enzyme target are of great importance because they will reduce the cost and time for therapy development.<sup>5-7</sup> Therefore, numerous computational approaches were advanced to carry out these tasks.<sup>8,9</sup> In terms of accuracy and required computing resources, these approaches can be roughly arranged into three groups: low accuracy and small consumption of central processing unit (CPU) time; medium in both accuracy and required CPU time; and accurate and precise approaches which require a large amount of computing resources. The first group involves

molecular docking<sup>10-13</sup> and quantitative structure-activity relationship (QSAR) approaches.<sup>10-15</sup> The second group includes fast pulling of ligand (FPL),<sup>16, 17</sup> umbrella sampling (US),<sup>3, 18, 19</sup> implicit ligand theory,<sup>20, 21</sup> linear interaction energy,<sup>22-25</sup> and molecular mechanism/Poisson-Boltzmann surface area (MM/PBSA),<sup>26-28</sup> approaches. The last group contains free energy perturbation (FEP),<sup>29, 30</sup> thermodynamics integration (TI),<sup>31, 32</sup> and non-equilibrium molecular dynamics simulations (NEMD).<sup>33, 34</sup> Moreover, the enhanced sampling techniques are also implemented into perturbation simulations in order to increase the obtained results.<sup>35-38</sup> However, these approaches would consume a huge of CPU time.

In order to calculate the binding free energy of a ligand to an enzyme, the molecular dynamics (MD) simulations were normally used to generate the equilibrium complex conformations, which conformations would be then employed as inputs into binding free energy investigations.<sup>39, 40</sup> However, before MD simulations were performed, because docking poses are required to use as MD initial conformations, molecular docking approaches are initially performed to preliminary estimate the ligand-binding pose and affinity.<sup>41, 42</sup> Moreover, molecular docking approaches have also been used for screening a large database of compounds,<sup>43, 44</sup> which can comprise several millions of compounds such as ZINC,<sup>45</sup> ChEMBL,<sup>46</sup> PubChem,<sup>47</sup> etc. Therefore, although their accuracy is not very high, molecular docking approaches play an important role in CADD.<sup>48</sup>

Autodock Vina (Vina)<sup>49</sup> is a free open-source application providing the ligand-binding affinity and pose rapidly.<sup>50</sup> Vina is broadly used in the scientific community with more than 10 000 citations since released in 2010 (Web of Knowledge). In particular, Vina was built with a completed empirical scoring function including the Gaussian steric interaction, repulsion, hydrogen bond, hydrophobic, and torsion metrics.<sup>49</sup> Moreover, the docking package is coded with the parallel computing ability, which makes Vina docking calculation run very fast.<sup>51</sup> Furthermore, user-friendly is also a strong point that attracts users of the package.<sup>52</sup> However, although Vina rapidly converges and adopts a large successful-docking rate, the correlation between docked and experimental data is low ( $R < 0.5$ ).<sup>51</sup> Due to the limitation, it is hard to rank the top-lead ligands based on the ligand-binding affinity. Therefore, in this work, the dependence of Vina performance on empirical parameters was assessed to search for optimal set of parameters which enhance docking accuracy. The task is of great interest, especially due to the widespread of Vina to estimate binding affinities and poses of various substrates to biomolecular targets.<sup>53-57</sup> In particular, 800 available ligands were redocked to their corresponding receptors using Vina with different empirical parameters. The list of complexes was reported in the previous study.<sup>51</sup> The dependence of Vina performance on individual empirical parameters was then clarified. Based on the observation, three sets of empirical parameters were proposed. The performance of the Vina with modified parameters on 800 complexes<sup>51</sup> was tested. Besides, 1315 available inhibitors were also redocked to the corresponding receptors in order to validate the obtained results. The optimized empirical parameters would improve the accuracy of Vina.

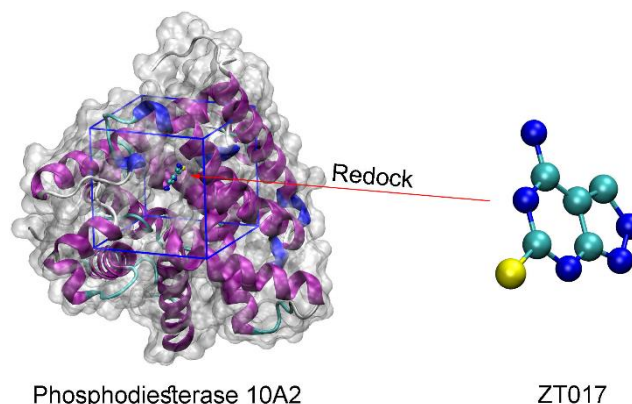
## MATERIALS & METHODS

### Complex Structures and Topologies

The complex conformations were download from the Protein Data Bank (PDB) according to the previous work<sup>51</sup> and also reported in detail in the Supporting Information. The topological PDBQT files for rigid receptors and flexible ligands were generated via AutodockTools with more details in the Supporting Information.<sup>11</sup> In particular, the Gasteiger-Marsili method was employed to calculate atomic charges.<sup>58, 59</sup>

### Molecular Docking Simulations

Vina was employed to redock several ligands to their corresponding receptors. In particular, the docking application was performed by using the globally searching exhaustiveness of 8, which corresponds to the default option. The maximum energy difference, which is the difference between the best and worst docking modes, was chosen as 7 kcal mol<sup>-1</sup>.<sup>51</sup> The docking grid was selected to be 20 × 20 × 20 Å, in which the grid center was the ligand center of mass (cf. **Figure 1**). Only one docking conformation corresponding to the lowest docked energy was recorded. Moreover, the AutoDock Vina 1.2 (Vina 1.2) was also performed with the same parameters, but the spacing was 0.375 Angstrom instead of 1.000 Angstrom in Vina.



**Figure 1.** Ligand was redocked to the binding site of the enzyme using Vina with various empirical parameters. The center of mass of the experimental binding ligand was used as the grid center using AutoDockTools.

### Structural Analysis

The root-mean-square deviation (RMSD) of non-hydrogen atoms between docked and experimental poses was determined using GROMACS tools.<sup>60</sup> The calculated error of correlation coefficient, successful-docking rate, and root-mean-square error (RMSE) was estimated by using 1,000 bootstrapping samples.<sup>61</sup> The computed error of ligand-binding free energy and RMSD was the standard error of the mean.

## RESULTS AND DISCUSSION

According to the previous assessment,<sup>51</sup> Vina rapidly converges since the accuracy insignificantly increased upon changing the globally searching exhaustiveness from 8 to 56 or 400, which corresponds to *short*, *medium*, and *long* options.<sup>51</sup> Moreover, it should be noted that increasing the

exhaustiveness causes at least ca. 7 times increase in the computing interval. It thus is hard to use *medium* or *long* options when screening a large database of compounds that consists of several thousand/million elements. Therefore, the default or the *short* docking option is widely used to complete the task. In this context, we redock available inhibitors to the corresponding targets using the *short* option as mentioned above. Furthermore, as mentioned above, Vina is an empirical approach, in which the application uses six parameters including *gauss1*, *gauss2*, *repulsion*, *hydrophobic*, *hydrogen bond*, and *rotation* terms to calculate the contribution of the Gaussian steric interaction, repulsion, hydrogen bond, hydrophobic, and torsion terms.<sup>49</sup> In order to assess Vina performance dependence upon empirical parameters, we changed the individual parameter with an among of 10 % of

each alteration (cf. **Table 1**). In particular, the four parameters *gauss1*, *repulsion*, *hydrophobic*, and *hydrogen bond* were thus changed in the range from -50 to +50 % with respect to the default values. The *gauss2*, and *rotation* parameters were varied in the range from -50 to +150 and -90 to +50 % of the parameters, respectively. Besides, although the docking of 800 ligands to corresponding receptors was previously completed using the original Vina, we also redocked these complexes via the unchanged docking application. Three new sets of empirical parameters were then proposed upon the understanding about the change of docking results via the alteration of empirical parameters, which named set1, set2, and set3. Therefore, 78 versions of Vina having different empirical parameters were compiled to redock ligands to receptors.

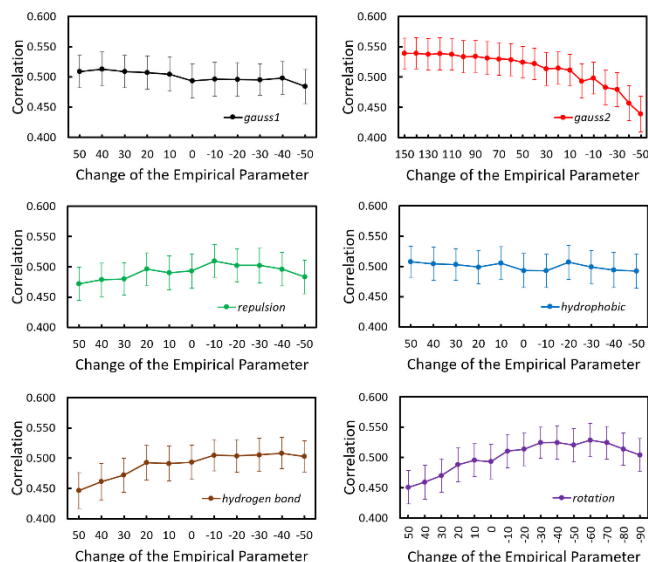
**Table 1.** The Different Empirical Parameter Values Was Tested.<sup>a</sup>

<i>NO</i>	<i>Change (%)</i>	<i>gauss1</i>	<i>gauss2</i>	<i>repulsion</i>	<i>hydrophobic</i>	<i>hydrogen bond</i>	<i>rotation</i>
1	-90						0.006431
2	-80						0.012861
3	-70						0.019292
4	-60						0.025722
5	-50	-0.017790	-0.002578	0.420123	-0.017535	-0.293720	0.029230
6	-40	-0.021347	-0.003094	0.504147	-0.021041	-0.352463	0.035076
7	-30	-0.024905	-0.003609	0.588172	-0.024548	-0.411207	0.040922
8	-20	-0.028463	-0.004125	0.672196	-0.028055	-0.469951	0.046768
9	-10	-0.032021	-0.00464	0.756221	-0.031562	-0.528695	0.052614
10	Default	-0.035579	-0.005156	0.840245	-0.035069	-0.587439	0.05846
11	+10	-0.039137	-0.005672	0.92427	-0.038576	-0.646183	0.064306
12	+20	-0.042695	-0.006187	1.008294	-0.042083	-0.704927	0.070152
13	+30	-0.046253	-0.006703	1.092319	-0.04559	-0.763671	0.075998
14	+40	-0.049811	-0.007218	1.176343	-0.049097	-0.822415	0.081844
15	+50	-0.053369	-0.007734	1.260368	-0.052604	-0.881159	0.087690
16	+60		-0.00825				
17	+70		-0.008765				
18	+80		-0.009281				
19	+90		-0.009796				
20	+100		-0.010312				
21	+110		-0.010828				
22	+120		-0.011343				
23	+130		-0.011859				
24	+140		-0.012374				
25	+150		-0.012890				

<sup>a</sup>Blank squares in the table mean that the changed parameters were not applied.

As mentioned above, the docking study using default empirical parameters was reproduced to compare with modified versions, in which the obtained results were mentioned in the Supporting Information 2. The Pearson correlation coefficient between docked and experimental values is of  $R_{\text{Default}} = 0.493 \pm 0.028$ , which is in good agreement with the previous work,  $R_{\text{Vina}}^{\text{short}} = 0.489 \pm 0.027$ .<sup>51</sup> The successful-docking rate which is defined as having RMSD less than 0.2 nm is of  $\hat{p} = 86 \pm 1 \%$ , which is larger than that reported in the previous reports using the *long* option,  $\hat{p} = 81 \%$ . It should be noted that the *long* option used the exhaustiveness of 400, which is required ca. 50 times longer computing time than the *short* option. The polar hydrogens were not automatically added to complexes in the previous work could be the cause of the smaller  $\hat{p}$ . However, the correlation coefficients were insignificantly changed. Besides, the average of binding energies over 800 complexes is of  $\Delta G_{\text{Default}} = -8.60 \pm 0.06 \text{ kcal mol}^{-1}$ , which is smaller than that from the previous study,  $\Delta G_{\text{Vina}}^{\text{short}} = -7.75 \pm 0.06$ .<sup>51</sup> The difference between theoretical and experimental results is of  $\delta = 0.62 \text{ kcal mol}^{-1}$ . The obtained outcome implies that the *hydrogen bond* parameter probably adopted a strong effect on docking pose and docking energy but less effect on docking accuracy.

The docking simulations using various empirical parameters, which were mentioned in **Table 1**, were performed. The docking results were described in the Supporting Information 2. Moreover, the accuracy of docking simulations was determined via the correlation coefficient analyses, in which the obtained values of  $R$  were shown in Table S1 of the Supporting Information. Furthermore, in the first steps, we have changed all of empirical parameters in the range from -50 to +50 % of their default values. The dependence of  $R$  upon the empirical parameters were shown in **Figure 2**. Interestingly, the  $R$  value is not sensitive to the change of the *gauss1* and *hydrophobic* parameters since it gives a relative deviation to the original one ( $R_{\text{Default}} = 0.493$ ) by amounts of 6 and 3 %, respectively. The  $R_{\text{gauss1}}$  reached the maximum value of  $0.513 \pm 0.028$  correspondings to the  $\text{gauss1} = -0.049811$  (+40%), the difference from the original Vina only is 4%. Besides, the alteration of the *hydrophobic* metric turns the  $R_{\text{hydrophobic}}$  ranging from 0.492 to 0.508. The largest value of  $R_{\text{hydrophobic}}$  thus differs the  $R_{\text{Default}}$  by an amount of 3%. Moreover, the obtained results suggested that the coefficient,  $R$ , is more sensitive with the change of the *repulsion* and *hydrogen bond* terms. In particular, the  $R_{\text{repulsion}}$  ranges from  $0.472 \pm 0.027$  to  $0.510 \pm 0.027$  and  $R_{\text{hydrogen bond}}$  ranges from  $0.447 \pm 0.030$  to  $0.508 \pm 0.026$  corresponding to relative variation of 8 and 12 %, respectively. However, the maximum value of  $R_{\text{repulsion}}$  and  $R_{\text{hydrogen bond}}$  only differ from the  $R_{\text{Default}}$  by 3%, respectively.

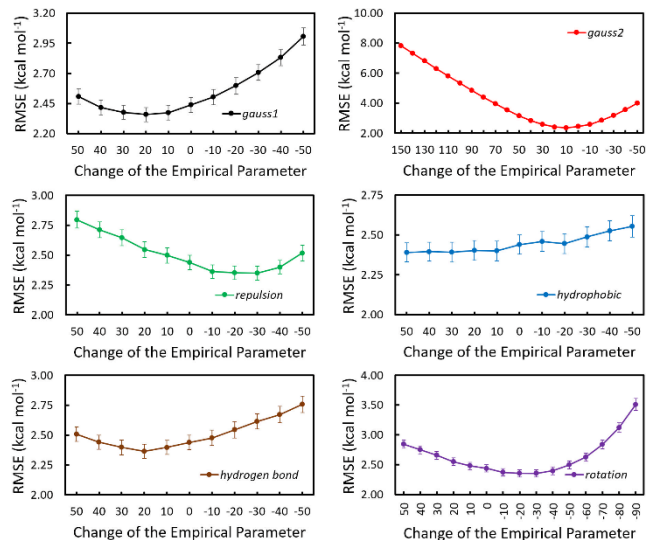


**Figure 2.** The correlation coefficient between docked and experimental binding free energy upon the changing of empirical parameters.

The story is significantly different when the alteration of the *gauss2* and *rotation* parameters were induced. These two terms are most influential to the Vina accuracy. When the changing from -50 to +50 % of the default value, the corresponding correlation coefficient  $R_{\text{gauss2}}$  and  $R_{\text{rotation}}$  range from  $0.439 \pm 0.030$  to  $0.524 \pm 0.024$  and  $0.451 \pm 0.027$  to  $0.525 \pm 0.028$ , which correspond to relative deviations of 17 and 15 %, respectively. The maximum values of  $R_{\text{gauss2}}$  and  $R_{\text{rotation}}$  differ from the  $R_{\text{Default}}$  with amounts of 6 and 6 %, respectively. Interestingly, although changing *gauss2* and *rotation* terms significantly increase the correlation coefficient, their influence are in the opposite direction. In particular, the  $R$  value tends to grow when the *rotation* parameter is decreased. On the other hand, the metric raises upon the increase of the *gauss2* term. Because the  $R$  values still increase upon the decreasing and increasing of the *gauss2* and *rotation* terms, respectively, the *gauss2* metric was increased to -0.012890 (+150%) of the parameter and the *rotation* value was decreased to 0.006431 (-90%) as mentioned in **Table 1**. The obtained  $R_{\text{gauss2}}$  and  $R_{\text{rotation}}$  depending the changes was also described in **Figure 2** and Table S1 of the Supporting Information. When the *rotation* parameter was gradually reduced, the  $R_{\text{rotation}}$  reached the maximum value,  $0.529 \pm 0.028$ , when the *rotation* value is of 0.025722 (-60%). Besides, the  $R_{\text{gauss2}}$  reached a maximum value of  $0.539 \pm 0.026$ , when the *gauss2* term was of -0.012374 (+140%). Overall, The obtained results are impressive since the maximum value of  $R_{\text{gauss2}}$  and  $R_{\text{rotation}}$  differ from the  $R_{\text{Default}}$  by amounts of 9 and 7 %, respectively.

The average of the docking energies among protein-ligand complexes was mentioned in Table S2 of the Supporting Information. In particular, the average value of  $\Delta G_{\text{Dock}}$  was linearly dependent on the change of empirical parameters. Especially, the average binding free energies are most sensitive to the change of the *gauss2* parameter, which is in good agreement with the observation that the  $R$  is mostly sensitive

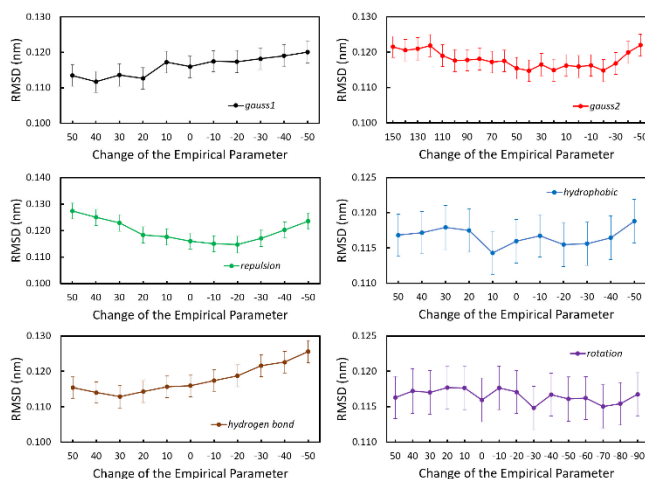
with the change of the *gauss2* parameter. The lowest  $\Delta G_{\text{Dock}} = -16.43 \pm 0.12$  kcal mol<sup>-1</sup> was observed when the *gauss2* parameter was assigned as -0.012890 (+150%). Besides, the highest  $\Delta G_{\text{Dock}} = -6.01 \pm 0.04$  kcal mol<sup>-1</sup> was obtained when the *gauss2* parameter was reduced by an amount of 50% (-0.002578) of the default parameter. Moreover, the difference between docked and experimental data is associated with the RMSE, which was mentioned in **Figure 3** and Table S3 of the Supporting Information. In particular, the RMSE was sensitive with the change of the *gauss1*, *gauss2* and *rotation* parameters, whereas the RMSE varies from 0.65 to 5.48 kcal mol<sup>-1</sup>. RMSE curves reached the minimum values, when the *gauss1*, *gauss2* and *rotation* metrics are of -0.042695 (+20%), -0.005672 (+10%) and 0.046768 (-20%), respectively.



**Figure 3.** RMSE of docking energies compared with the respective experiments. The results were obtained when the empirical parameters were altered. The computed error was estimated using 1000 rounds of the bootstrapping analysis.

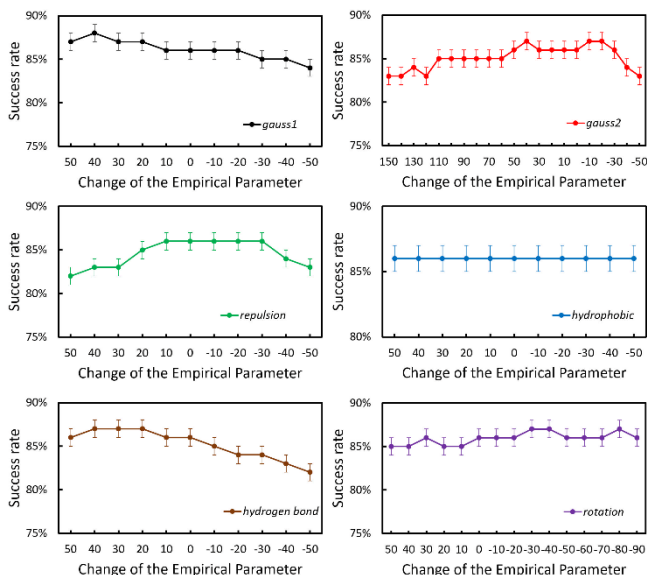
As mentioned above, the binding poses of ligands to receptors via molecular docking simulations also play an important factor since they would be used as initial conformations of MD-refined simulations.<sup>62-64</sup> Because the obtained results of the calculated binding free energy via MD simulations are sensitive to the difference of binding conformations of complexes.<sup>65</sup> Indeed, the computational binding structure is more fitting to the native binding shape meaning that the estimating binding free energy is more accurate. Consequently, the better docking pose is the shape that is more fitting to the experimental binding conformation.<sup>66</sup> The MD-refined simulations would be faster to reach the stabilized conformations. The consumption of computing resources would be thus reduced. Therefore, the successful-docking rate,  $\hat{p}$ , was carefully investigated, in which the RMSD of non-hydrogen atoms between docked and experimental structures were assessed (cf. **Figure 4** and Table S4 of the Supporting Information). Normally, a successful-docked shape is the docked structure having RMSD to the respective experimental shape is less than 0.2 nm.<sup>67</sup> However,

in this work, we also assessed the successful-docking rate with a RMSD cutoff of 0.15, 0.10, and 0.05 nm.<sup>51</sup>



**Figure 4.** The dependence of RMSD upon the change of the empirical parameters. The computed error is the standard error of the mean.

The original Vina adopted a mean RMSD of  $0.116 \pm 0.003$  nm compared to the respective experiments. The successful-docking rate with an RMSD cutoff of 0.20 nm is of  $\hat{p} = 86 \pm 1$  %. Besides, the modified Vina formed a mean RMSD ranging from  $0.112 \pm 0.003$  to  $0.127 \pm 0.003$  nm, resulting in adopting  $\hat{p}$  value in the range from  $82 \pm 1$  to  $88 \pm 1$  % corresponding with the *repulsion* and *gauss1* terms of 1.260368 (+50%) and -0.049811 (+40%), respectively. In particular, the dependence on alteration in the empirical parameters of  $\hat{p}$  with an RMSD cutoff off 0.20 nm was fully reported in **Figure 5** and Table S5 of the Supporting Information. It should be noted that the obtained mean values of RMSD well correlated with the successful-docking rate. The smaller the mean RMSD means the larger the success rate of docking simulations. Moreover,  $\hat{p}$  reached its largest value of 6%, when the *hydrogen bond* term was altered within their respective ranges. The observation are in good agreement with discussion above. The  $\hat{p}_{\text{hydrogen bond}}$  value reached the largest values when the *hydrogen bond* metric is of -0.763671 (+30%). The  $\hat{p}$  is not sensitive to the change of the *hydrophobic* parameter since the value varies within a range of <1% only. Furthermore, changing the *gauss1* term provides the largest successful-docking rate  $\hat{p}_{\text{gauss1}}$  of  $88 \pm 1$  %. The alteration of *repulsion* and *hydrophobic* terms are not enlarge the  $R$  value. Other parameters including *gauss2*, *hydrogen bond*, and *rotation* metrics only formed the maximum value  $\hat{p} = 87 \pm 1$  %. Furthermore, when the other cutoff was applied including 0.15, 0.10, and 0.05 nm, the  $\hat{p}$  metric became more sensitive to the *gauss1*, *gauss2*, *repulsion*, and *hydrogen bond* terms and insensitive with the *hydrophobic* and *rotation* terms (cf. Tables S6-S8 and Figures S2-S4 of the Supporting Information). Consequently, the *repulsion* metric dominates over all of the metrics in influencing the successful-docking rate, which  $\hat{p}$  was changed in the range from 4 to 6 % upon the modification of the *repulsion* term.



**Figure 5.** Changing of the successful-docking rate,  $\hat{p}$ , upon the alteration of the empirical parameters with an RMSD cutoff of 0.20 nm. The computed error was calculated via 1000 rounds of the bootstrapping analysis.

Considering the overall influence of empirical parameters on the Vina results, we proposed three sets of empirical parameters, which were shown in **Table 2**. In particular, over three sets of parameters, the *gauss1* term was selected as -0.049811 (+40%) since both  $R_{gauss1}$  and  $\hat{p}$  reached the maximum values with the corresponding parameters compared the original one (**Figure 5**). The *gauss2* term was chosen as -0.007218 (+40%) for set1 and set2 of parameters, because the  $R_{gauss2}$  formed an appropriate value, which is larger than the original one by an amount of 6%, and the  $\hat{p}$  term achieved the largest amount of 87%. In set3, the *gauss2* parameter was selected as -0.007734 (+50%) since the obtained  $R_{gauss2}$  and  $\hat{p}_{gauss2}$  are almost the same as the corresponding values of set1 and set2. Moreover, the *repulsion* parameter was of 0.756221 (-10%) for set1 and set2 of empirical parameters, because the modified Vina adopted the largest correlation coefficient ( $0.510 \pm 0.027$ ). In set3, the *repulsion* term was picked as 0.672196 (-20%) since the  $\hat{p}$  value reached the largest amount (**Figure 5**). The *hydrophobic* term was -0.028055 (-20%) for three sets since the docking package formed the maximum value of  $R_{hydrophobic} = 0.507 \pm 0.028$ . Besides, interestingly, the  $\hat{p}_{hydrophobic}$  is slightly increased corresponding to this *hydrophobic* term (**Figure 5**). Furthermore, in set1 and set2, the *hydrogen bond* metric was changed to -0.352463 (-40%) due to forming the strongest correlation to the experiments. However, in set3, the *hydrogen bond* value was chosen as -0.528695 (-10%), because the  $R$  value was increased but the  $\hat{p}$  metric was insignificantly decreased (cf. **Figure 2** and **Figure 5**). Finally, in set1 of empirical parameters, the *rotation* term was 0.025722 (-60%), because the accuracy of the result was maximized with a value of  $R_{rotation} = 0.529 \pm 0.028$  (**Figure 2**). Besides, the *rotation* parameter of set2 and set3 was selected as 0.012861 (-80%) since the modified Vina formed the largest successful-docking rate,  $87 \pm 1\%$ , and appropriate correlation coefficient,  $0.514 \pm 0.027$ , (**Figure 2**).

**Table 2.** Proposed Sets of Empirical Parameters.

Nº	Parameter	Set1	Set2	Set3	Default <sup>a</sup>
1	<i>gauss1</i>	-0.049811	-0.049811	-0.049811	-0.035579
2	<i>gauss2</i>	-0.007218	-0.007218	-0.007734	-0.005156
3	<i>repulsion</i>	0.756221	0.756221	0.672196	0.840245
4	<i>hydrophobic</i>	-0.031562	-0.031562	-0.031562	-0.035069
5	<i>hydrogen bond</i>	-0.469951	-0.469951	-0.528695	-0.587439
6	<i>rotation</i>	0.025722	0.012861	0.012861	0.058460

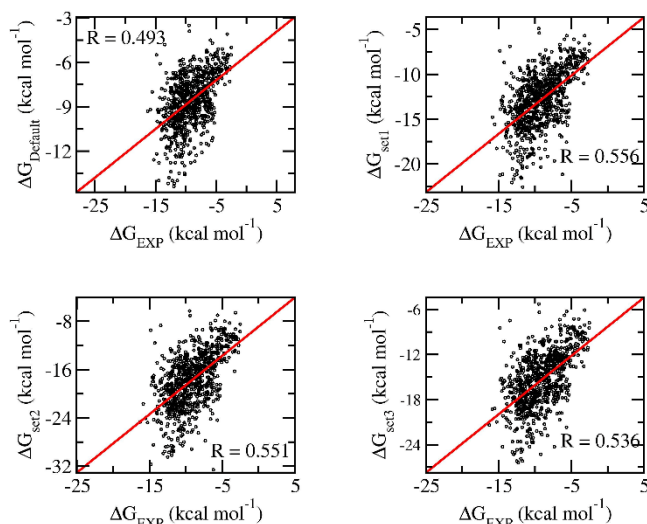
<sup>a</sup>The default empirical parameters were reported in the previous study.<sup>49</sup>

Molecular docking simulations used three sets of empirical parameters were performed over 800 complexes, which complexes were reported in the previous work.<sup>51</sup> The obtained results are displayed in **Table 3** and **Figure 6**. The detailed results were shown in Supporting Information 2. Interestingly, according to the selection of empirical parameters, the increase of the correlation coefficient depends on set1, set2, and set3, in which the obtained coefficients are of  $R_{set1} = 0.556 \pm 0.025$ ,  $R_{set2} = 0.551 \pm 0.026$ , and  $R_{set3} = 0.536 \pm 0.028$ , respectively. The obtained correlations are significantly larger than that of the original Vina ( $R_{Default} = 0.493 \pm 0.028$ ). Moreover, the performance of AutoDock Vina 1.2 (Vina 1.2),<sup>68</sup> a new version of Vina supporting the AutoDock4.2 scoring function, was also assessed and reported in **Table 3** and the Supporting Information 2. Although the Vina 1.2,  $R_{Vina\ 1.2} = 0.503 \pm 0.029$ , is a more accurate ranking than the original Vina,  $R_{Default} = 0.493 \pm 0.028$ , the modified Vina formed a larger correlation coefficient (**Table 3**). Furthermore, the modified Vina formed a larger binding affinity compared with the original one. Furthermore, because, as mentioned above, the modified parameters were chosen by considering to increase the accuracy, the obtained  $\hat{p}$  values of modified packages were slightly decreased (cf. **Table 3**).

**Table 3.** Calculated Metrics of the Modified Vina in Comparison with the Original Version.<sup>a</sup>

Nº	Package	$R$	$\Delta G_{Dock}$	RMSD	$\hat{p}$
1	set1	$0.556 \pm 0.025$	$-12.88 \pm 0.11$	$0.119 \pm 0.003$	$84 \pm 1$
2	set2	$0.551 \pm 0.026$	$-17.77 \pm 0.17$	$0.120 \pm 0.003$	$83 \pm 1$
3	set3	$0.536 \pm 0.028$	$-15.44 \pm 0.14$	$0.118 \pm 0.003$	$85 \pm 1$
4	Default	$0.493 \pm 0.028$	$-8.60 \pm 0.06$	$0.116 \pm 0.003$	$86 \pm 1$
5	Vina 1.2	$0.503 \pm 0.029$	$-8.60 \pm 0.06$	$0.115 \pm 0.003$	$87 \pm 1$

<sup>a</sup>The unit of  $\Delta G_{Dock}$  and RMSD is kcal mol<sup>-1</sup> and nm, respectively. The computed error of  $\Delta G_{Dock}$  and RMSD is standard error of the mean. The  $\hat{p}$  value was calculated within a cutoff 0.2 nm.



**Figure 6.** Correlation coefficients between original/modified Vina versus experiments. The results were obtained over 800 complexes, which were listed in the previous work.<sup>51</sup>

The influence of modified empirical parameters on individual enzymic targets was also evaluated and the results were shown in **Table 4**. Vina with the default parameters formed appropriate correlation coefficients ( $R_{\text{Default}} \geq 0.500$ ) over 31 targets. Vina with set1 parameters formed more suitable results with 32 targets having appropriate coefficients ( $R_{\text{set1}} \geq 0.500$ ). Besides, Vina with set2 and set3 parameters adopted suitable correlation coefficients for 33 targets. Absolutely, although the new sets of empirical parameters formed overestimated results, a better correlation to the respective experiments was observed. It is a critical increase because when screening a large number of ligands, we first need to know the ranking of ligand-binding affinity.

**Table 4.** The Correlation Coefficient between the Docking and Experimental Binding Free Energy for 48 receptors.

N <sup>o</sup>	Complexes	$R_{\text{Default}}$	$R_{\text{set1}}$	$R_{\text{set2}}$	$R_{\text{set3}}$
1	3-Dehydroquinase Dehydratase	0.708	0.651	0.657	0.671
2	Alpha-L-Fucosidase	0.920	0.897	0.903	0.935
3	AmpC Beta-Lactamase	0.686	0.711	0.735	0.760
4	Beta-Lactamase CTX-M-9a	0.951	0.918	0.927	0.965
5	Beta-Galactosidase	0.828	0.819	0.758	0.813
6	Beta-Hexosaminidase	-0.452	0.440	0.571	0.496
7	Calmodulin-Domain Protein Kinase 1	0.785	0.800	0.786	0.786
8	Carboxypeptidase A	0.823	0.922	0.937	0.929
9	Catechol O-methyltransferase	0.463	0.694	0.714	0.627
10	Cathepsin K	0.428	0.441	0.428	0.298
11	Coagulation Factor X	0.587	0.678	0.737	0.734
12	Dipeptidyl-Peptidase 4	0.299	0.212	0.216	0.236
13	Estrogen Receptor	0.459	0.568	0.544	0.555
14	Glutamate Receptor, Ionotropic Kainate 1	0.754	0.696	0.650	0.667
15	HSP90	0.236	0.338	0.399	0.366
16	M1 Family Aminopeptidase	0.660	0.758	0.758	0.757
17	Macrophage metalloelastase	0.867	0.870	0.846	0.846
18	Mitogen-activated protein kinase 1	0.490	0.644	0.673	0.693
19	Neuraminidase	0.439	0.227	0.201	0.217
20	N-terminal Human Maltase-Glucoamylase	0.620	0.353	0.337	0.422
21	Phosphodiesterase 10A2	0.733	0.782	0.786	0.796
22	Purine nucleoside phosphorylase	0.835	0.742	0.722	0.741
23	Queuine tRNA-ribosyltransferase	0.664	0.669	0.668	0.686
24	Scytalone Dehydratase	0.576	0.516	0.511	0.525
25	Serine/Threonine-Protein Kinase Chk1	0.840	0.841	0.776	0.819
26	Thermolysin	0.636	0.781	0.777	0.838
27	Thrombin	0.084	0.233	0.256	0.279
28	Thymidylate Kinase	-0.001	0.007	0.017	-0.052
29	Thymidylate Synthase	0.539	0.438	0.422	0.353
30	Trypsin	0.671	0.617	0.598	0.611
31	Tyrosine Phosphatase 1b	0.222	0.525	0.601	0.636
32	Tyrosine-protein kinase JAK1	0.597	0.754	0.820	0.797
33	Tyrosine-protein kinase JAK2	0.594	0.728	0.728	0.784
34	Urokinase-Type Plasminogen Activator	0.553	0.758	0.780	0.796
35	Beta-secretase 1	0.607	0.623	0.580	0.587
36	Carbonic anhydrase 2	0.041	0.101	0.074	0.039
37	Cell Division Protein Kinase 2	0.729	0.749	0.762	0.776
38	Cyclin Dependent Kinase 2	0.637	0.643	0.657	0.654
39	Dehydrosqualene synthase	0.515	0.449	0.456	0.556
40	Endothiapepsin	0.210	0.235	0.112	0.161
41	Factor XA	-0.184	-0.215	-0.208	-0.206
42	Glycogen phosphorylase	0.535	0.419	0.352	0.432
43	NS3 protease, NS4A protein	0.218	0.459	0.475	0.441
44	Penicillin Amidohydrolase	-0.569	-0.611	-0.687	-0.744
45	Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit gamma isoform	0.742	0.787	0.791	0.767
46	Ribonuclease A	0.471	0.611	0.610	0.663
47	Stromelysin-1	0.662	0.815	0.852	0.877
48	Alpha-Mannosidase 2	0.918	0.717	0.696	0.728

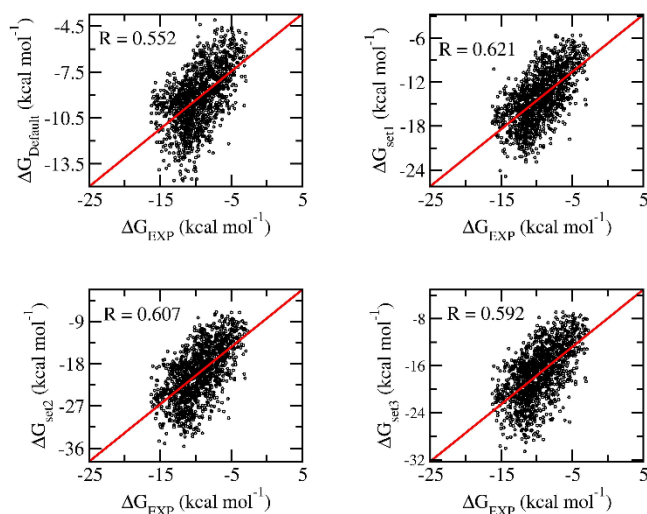
In addition, a further validated investigation about the performance of the modified Vina with the difference of empirical parameters was performed in comparison with the original one. 1315 complexes from the version 2019 of PDBbind refined structures, which were mentioned in Supporting Information, were redocked by using the original and

modified Vina package. The obtained results were mentioned in **Table 5** and **Figure 7**. All of the modified versions of Vina formed larger correlation coefficients compared with the original one ( $R_{\text{Default}} = 0.552 \pm 0.018$ ) and the Vina 1.2 version ( $R_{\text{Vina 1.2}} = 0.549 \pm 0.017$ ). The obtained  $R$  of modified Vina diffuse in the range from  $0.592 \pm 0.017$  to  $0.621 \pm 0.016$ . The Vina with set1 parameters adopted the largest coefficient of  $R_{\text{set1}} = 0.621 \pm 0.016$ . Moreover, the successful-docking rates were reduced, which range from  $80 \pm 1$  to  $82 \pm 1$  % in comparison with the default package,  $\hat{p}_{\text{Default}} = 84 \pm 1$  %, and the 1.2 version,  $\hat{p}_{\text{Vina 1.2}} = 83 \pm 1$  %.. The original Vina adopted the larger value  $\hat{p}_{\text{Default}}$  compared to the modified packages (cf. **Table 5**) possibly because the version 2009 of PDBbind refined complexes were used to train the empirical parameters of the original Vina.<sup>49</sup> Fortunately, the set1 parameters formed a slight decrease of  $\hat{p}$  term ( $\hat{p}_{\text{set1}} = 82 \pm 1$  %). However, reaching a higher correlation coefficient is the most important since docking simulations were mainly used to relatively rank the ligand-binding affinity (as mentioned above). Furthermore, the modified Vina adopted rigidly overestimate results. Although it is not an important issue since docking simulations were mainly used to relatively rank the ligand-binding affinity, Vina with set1 parameters formed the smallest value of the difference of docking energy to experimental value compared with the other modified Vina versions (**Table 5**). Overall, we may assume that set1 is the most appropriate since it stably forms the largest  $R$  and appropriate  $\hat{p}$  over various groups of complexes.

**Table 5.** Calculated Results of the Modified Vina Compared with the Original Version.<sup>a</sup>

N <sup>o</sup>	Package	$R$	$\Delta G_{\text{Dock}}$	$\text{RMSD}$	$\hat{p}$
1	set1	$0.621 \pm 0.016$	$-14.20 \pm 0.10$	$0.127 \pm 0.003$	$82 \pm 1$
2	set2	$0.607 \pm 0.016$	$-20.04 \pm 0.16$	$0.131 \pm 0.003$	$80 \pm 1$
3	set3	$0.592 \pm 0.017$	$-17.31 \pm 0.13$	$0.124 \pm 0.003$	$82 \pm 1$
4	Default	$0.552 \pm 0.018$	$-9.20 \pm 0.05$	$0.116 \pm 0.003$	$84 \pm 1$
5	Vina 1.2	$0.549 \pm 0.017$	$-9.15 \pm 0.05$	$0.118 \pm 0.003$	$83 \pm 1$

<sup>a</sup>The unit of  $\Delta G_{\text{Dock}}$  and RMSD is kcal mol<sup>-1</sup> and nm, respectively. The computed error of  $\Delta G_{\text{Dock}}$  and RMSD is standard error of the mean. The  $\hat{p}$  value was calculated within a cutoff 0.2 nm.



**Figure 7.** Correlation coefficients between original/modified Vina versus experiments. The results were obtained over 1315 complexes from PDB refine structures.

## CONCLUSIONS

Autodock Vina is a rapidly convergence molecular docking approach as well as forms a large successful-docking rate. However, adopting low correlated results to the respective experiments is the scheme weaknesses. Normally, docking simulations were mostly employed to relatively rank the ligand-binding affinity, the obtained accuracy is thus more important than the other metrics. In this work, we have evaluated the dependence of docking results over the change of empirical parameters. Although changing of six parameters alters the obtained correlation coefficient  $R$ , the *gauss2* and *rotation* terms form more effects. Moreover, the successful-docking rate  $\hat{p}$  are sensitive with the alteration of the *gauss1*, *gauss2*, and *repulsion* parameters.

Three sets of empirical parameters were proposed as well as set1 (more priority for accuracy), set2 (keep a balance between  $R$  and  $\hat{p}$ ), and set3 (more priority for  $\hat{p}$ ) based on the knowledge on the dependence of Vina on individual empirical parameters. The testing study of three modified Vina over 800 complexes was then carried out. The obtained correlation coefficients were significantly larger than that by the original and 1.2 versions. Therefore, the Vina with proposed sets of parameters can provide more accurate results. Moreover, all of the modified versions formed appropriate correlation coefficients ( $R > 0.500$ ) for  $\geq 32$  targets, where the corresponding number provided by the original Vina is 31 only. Therefore, the Vina with proposed sets of parameters can apply wider compared with the default parameters.

Validating investigations over 1315 complexes from the version 2019 of PDBbind refined structures were also performed. The obtained correlation of three modified Vina was significantly larger than the original one suggesting that the docking approach with proposed parameters can rank ligand-binding affinity with more accuracy. However, because of the decrease of the successful-docking rate, we have suggested that set1 and set2 parameters are more appropriate than set3 parameters.

## ASSOCIATED CONTENT

### Data and Software Availability

Information need to reproduce results in the article are reported in the main text and the Supporting Information and Supporting Information 2.

### Supporting Information

Supporting Information Available: Information about how to prepare PDBQT file for receptors and ligands. The correlation between docked and experimental ligand-binding affinity. The average of docking energy was provided by molecular docking simulations. The RMSE between calculated and experimental data. The RMSD of non-hydrogen atoms between calculated and experimental binding poses. The successful-docking rate of the docking results with an RMSD cutoff of 0.20, 0.15, 0.10, and 0.05 nm. Python code used to calculate correlation coefficient, RMSE, successful-docking rate, and their computed error via bootstrapping analysis. Besides, the list of complexes and docking results in the Supporting Information 2. The material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*Email: [ngosontung@tdtu.edu.vn](mailto:ngosontung@tdtu.edu.vn)

### Author Contributions

All authors designed the studies, collected, analyzed data, and wrote the manuscript.

### Notes

The authors declare no competing financial interests.

## ACKNOWLEDGMENT

This research was funded by Foundation for Science and Technology Development of Ton Duc Thang University (FOSTECT), website: <http://fostect.tdtu.edu.vn>, under Grant FOSTECT.2019B.08.

## REFERENCES

1. Ryde, U.; Soderhjelm, P., Ligand-Binding Affinity Estimates Supported by Quantum-Mechanical Methods. *Chem. Rev.* **2016**, *116* (9), 5520-5566.
2. Jannat, S.; Balupuri, A.; Ali, M. Y.; Hong, S. S.; Choi, C. W.; Choi, Y.-H.; Ku, J.-M.; Kim, W. J.; Leem, J. Y.; Kim, J. E.; Shrestha, A. C.; Ham, H. N.; Lee, K.-H.; Kim, D. M.; Kang, N. S.; Park, G. H., Inhibition of  $\beta$ -site amyloid precursor protein cleaving enzyme 1 and cholinesterases by pterosins via a specific structure-activity relationship with a strong BBB permeability. *Exp. Mol. Med.* **2019**, *51* (2), 12.
3. Ngo, S. T., Estimating the ligand-binding affinity via  $\lambda$ -dependent umbrella sampling simulations. *J. Comput Chem* **2021**, *42*, 117-123.
4. Macchiagodena, M.; Pagliai, M.; Karrenbrock, M.; Guarnieri, G.; Iannone, F.; Procacci, P., Virtual Double-System Single-Box: A Nonequilibrium Alchemical Technique for Absolute Binding Free Energy Calculations: Application to Ligands of the SARS-CoV-2 Main Protease. *J. Chem. Theor. Comput.* **2020**, *16* (11), 7160-7172.
5. Yu, W.; MacKerell, A. D., Computer-Aided Drug Design Methods. In *Antibiotics: Methods and Protocols*, Sass, P., Ed. Springer New York: New York, NY, 2017; pp 85-106.
6. Nguyen, T. H.; Zhou, H.-X.; Minh, D. D. L., Using the Fast Fourier Transform in Binding Free Energy Calculations. *J. Comput. Chem.* **2018**, *39* (11), 621-636.
7. Marshall, G. R., Computer-Aided Drug Design. *Ann. Rev. Pharmacol. Toxicol.* **1987**, *27*, 193-213.
8. Decherchi, S.; Cavalli, A., Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. *Chem. Rev.* **2020**.
9. Limongelli, V., Ligand Binding Free Energy and Kinetics Calculation in 2020. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10* (4), e1455.
10. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R., Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267* (3), 727-748.
11. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785-2791.
12. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem* **2004**, *47* (7), 1739-1749.
13. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem* **2004**, *47* (7), 1750-1759.
14. Holliday, J. D.; Ranade, S. S.; Willett, P., A Fast Algorithm For Selecting Sets Of Dissimilar Molecules From Large Chemical Databases. *Quant Struct-Act Rel* **1995**, *14* (6), 501-506.
15. Gil-Redondo, R.; Klett, J.; Gago, F.; Morreale, A., gCOMBINE: A Graphical User Interface to Perform Structure-Based Comparative Binding Energy (COMBINE) Analysis on a Set of Ligand-Receptor Complexes. *Proteins* **2010**, *78* (1), 162-172.
16. Ngo, S. T.; Hung, H. M.; Nguyen, M. T., Fast and Accurate Determination of the Relative Binding Affinities of Small Compounds to HIV-1 Protease using Non-Equilibrium Work. *J. Comput. Chem.* **2016**, *37* (31), 2734-2742.
17. Ngo, S. T.; Nguyen, M. T.; Nguyen, M. T., Determination of the absolute binding free energies of HIV-1 protease inhibitors using non-equilibrium molecular dynamics simulations. *Chem. Phys. Lett.* **2017**, *676*, 12-17.
18. Ngo, S. T.; Vu, K. B.; Bui, L. M.; Vu, V. V., Effective Estimation of Ligand-Binding Affinity Using Biased Sampling Method. *ACS Omega* **2019**, *4* (2), 3887-3893.
19. Lan, N. T.; Vu, K. B.; Dao Ngoc, M. K.; Tran, P.-T.; Hiep, D. M.; Tung, N. T.; Ngo, S. T., Prediction of AChE-ligand affinity using the umbrella sampling simulation. *J Mol Graph Model* **2019**, *93*, 107441.
20. Nguyen, T. H.; Minh, D. D. L., Implicit Ligand Theory for Relative Binding Free Energies. *J. Chem. Phys.* **2018**, *148* (10), 104114.

21. Nguyen, T. H.; Minh, D. D. L., Implicit Ligand Theory for Relative Binding Free Energies: II. An Estimator Based on Control Variates. *J. Phys. Commun.* **2020**, *4* (11), 115010.
22. Hansson, T.; Marelus, J.; Åqvist, J., Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aid. Mol. Des.* **1998**, *12* (1), 27-35.
23. Gutiérrez-de-Terán, H.; Åqvist, J., Linear Interaction Energy: Method and Applications in Drug Design. In *Computational Drug Discovery and Design*, Baron, R., Ed. Springer New York: 2012; Vol. 819, pp 305-323.
24. Nunes-Alves, A.; Arantes, G. M., Ligand-Receptor Affinities Computed by an Adapted Linear Interaction Model for Continuum Electrostatics and by Protein Conformational Averaging. *J. Chem. Inf. Model.* **2014**, *54* (8), 2309-2319.
25. Ngo, S. T.; Hong, N. D.; Quynh Anh, L. H.; Hiep, D. M.; Tung, N. T., Effective estimation of the inhibitor affinity of HIV-1 protease via a modified LIE approach. *RSC Adv* **2020**, *10* (13), 7732-7739.
26. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33* (12), 889-897.
27. Kuhn, B.; Kollman, P. A., Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, *43* (20), 3786-3791.
28. Wang, W.; Kollman, P. A., Computational study of protein specificity: the molecular basis of HIV-1 protease drug resistance. *Proc. Natl. Acad. Sci. USA* **2001**, *98* (26), 14937-14942.
29. Zwanzig, R. W., High-temperature equation of state by a perturbation method. I. nonpolar gases. *J. Chem. Phys.* **1954**, *22* (8), 1420-1426.
30. Helms, V.; Wade, R. C., Computational Alchemy To Calculate Absolute Protein-Ligand Binding Free Energy. *J. Am. Chem. Soc.* **1998**, *120* (12), 2710-2713.
31. Kirkwood, J. G., Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3* (5), 300-313.
32. Kollman, P., Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93* (7), 2395-2417.
33. Jarzynski, C., Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys Rev E* **1997**, *56* (5), 5018-5035.
34. Ytreberg, F. M., Absolute FKBP Binding Affinities Obtained via Non-Equilibrium Unbinding Simulations. *J. Chem. Phys.* **2009**, *130* (16), 164906.
35. Ngo, S. T.; Nguyen, T. H.; Tung, N. T.; Nam, P. C.; Vu, K. B.; Vu, V. V., Oversampling Free Energy Perturbation Simulation in Determination of the Ligand-Binding Free Energy. *J. Comput. Chem* **2020**, *41* (7), 611-618.
36. Jiang, W.; Roux, B., Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2010**, *6* (9), 2559-2565.
37. Meli, M.; Colombo, G., A Hamiltonian Replica Exchange Molecular Dynamics (MD) Method for the Study of Folding, Based on the Analysis of the Stabilization Determinants of Proteins. *International Journal of Molecular Sciences* **2013**, *14* (6), 12157-12169.
38. Jiang, W.; Thirman, J.; Jo, S.; Roux, B., Reduced Free Energy Perturbation/Hamiltonian Replica Exchange Molecular Dynamics Method with Unbiased Alchemical Thermodynamic Axis. *J. Phys. Chem. B* **2018**, *122* (41), 9435-9442.
39. Ngo, S. T.; Hung Minh, N.; Le Thi Thuy, H.; Pham Minh, Q.; Vi Khanh, T.; Nguyen Thanh, T.; Van, V., Assessing Potential Inhibitors for SARS-CoV-2 Main Protease from Available Drugs using Free Energy Perturbation Simulations. *RSC Adv* **2020**, *10*, 40284-40290.
40. Ngo, S. T.; Quynh Anh Pham, N.; Thi Le, L.; Pham, D.-H.; Vu, V. V., Computational Determination of Potential Inhibitors of SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **2020**, *60* (12), 5771-5780.
41. Tam, N. M.; Pham, M. Q.; Ha, N. X.; Nam, P. C.; Phung, H. T. T., Computational Estimation of Potential Inhibitors from Known Drugs Against the Main Protease of SARS-CoV-2. *RSC Adv* **2021**, *11* (28), 17478-17486.
42. Pham, M. Q.; Vu, K. B.; Han Pham, T. N.; Thuy Huong, L. T.; Tran, L. H.; Tung, N. T.; Vu, V. V.; Nguyen, T. H.; Ngo, S. T., Rapid prediction of possible inhibitors for SARS-CoV-2 main protease using docking and FPL simulations. *RSC Adv* **2020**, *10* (53), 31991-31996.
43. Shukla, R.; Munjal, N. S.; Singh, T. R., Identification of Novel Small Molecules Against GSK3 $\beta$  for Alzheimer's Disease using Chemoinformatics Approach. *J. Mol. Graph. Modell.* **2019**, *91*, 91-104.
44. Szaszko, M.; Hajdú, I.; Flachner, B.; Dobi, K.; Magyar, C.; Simon, I.; Lőrincz, Z.; Kapui, Z.; Pázmány, T.; Cseh, S.; Dormán, G., Identification of potential glutamyl cyclase inhibitors from lead-like libraries by in silico and in vitro fragment-based screening. *Mol. Divers.* **2017**, *21* (1), 175-186.
45. Sterling, T.; Irwin, J. J., ZINC 15--Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55* (11), 2324-2337.
46. Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, María P.; Mosquera, Juan F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, Chris J.; Segura-Cabrera, A.; Hersey, A.; Leach, Andrew R., ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* **2018**, *47* (D1), D930-D940.
47. Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H., PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44* (D1), D1202-D1213.
48. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W., Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66* (1), 334-395.
49. Trott, O.; Olson, A. J., Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455-461.
50. Gaillard, T., Evaluation of AutoDock and AutoDock Vina on the CASF-2013 Benchmark. *J. Chem. Inf. Model.* **2018**, *58* (8), 1697-1706.
51. Nguyen, N. T.; Nguyen, T. H.; Pham, T. N. H.; Huy, N. T.; Bay, M. V.; Pham, M. Q.; Nam, P. C.; Vu, V. V.; Ngo, S. T., Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity. *J. Chem. Inf. Model.* **2020**, *60* (1), 204-211.
52. Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J., Computational protein-ligand

- docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **2016**, *11*, 905.
53. Noike, M.; Matsui, T.; Ooya, K.; Sasaki, I.; Ohtaki, S.; Hamano, Y.; Maruyama, C.; Ishikawa, J.; Satoh, Y.; Ito, H.; Morita, H.; Dai, T., A peptide ligase and the ribosome cooperate to synthesize the peptide phegantomycin. *Nat. Chem. Biol.* **2014**, *11*, 71.
  54. Grither, W. R.; Longmore, G. D., Inhibition of tumor-microenvironment interaction and tumor invasion by small-molecule allosteric inhibitor of DDR2 extracellular domain. *Proc Natl Acad Sci U S A* **2018**, *115* (33), E7786.
  55. Vu, V. V.; Hangasky, J. A.; Detomasi, T. C.; Henry, S. J. W.; Ngo, S. T.; Span, E. A.; Marletta, M. A., Substrate selectivity in starch polysaccharide monooxygenases. *J. Biol. Chem.* **2019**, *294*, 12157-12166.
  56. Caffalette, C. A.; Corey, R. A.; Sansom, M. S. P.; Stansfeld, P. J.; Zimmer, J., A lipid gating mechanism for the channel-forming O antigen ABC transporter. *Nat Comm* **2019**, *10* (1), 824.
  57. Ngo, S. T.; Tran-Le, P. D.; Ho, G. T.; Le, L. Q.; Bui, L. M.; Vu, B. K.; Thu Phung, H. T.; Nguyen, H.-D.; Vo, T.-S.; Vu, V. V., Interaction of carbohydrate binding module 20 with starch substrates. *RSC Adv* **2019**, *9* (43), 24833-24842.
  58. Gasteiger, J.; Marsili, M., New Model for Calculating Atomic Charges in Molecules. *Tetrahedron Lett* **1978**, *19*.
  59. Gasteiger, J.; Marsili, M., Iterative Partial Equalization of Orbital Electronegativity—A Rapid Cccess to Atomic Charges. *Tetrahedron* **1980**, *36* (22), 3219-3228.
  60. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.
  61. Efron, B., Bootstrap Methods: Another Kook at the Jackknife. *Ann. Stat.* **1979**, *7*, 1-26.
  62. Tam, N. M.; Nam, P. C.; Quang, D. T.; Tung, N. T.; Vu, V. V.; Ngo, S. T., Binding of Inhibitors to the Monomeric and Dimeric SARS-CoV-2 Mpro. *RSC Adv* **2021**, *11*, 2926-2934.
  63. Khanal, P.; Dey, Y. N.; Patil, R.; Chikhale, R.; Wanjari, M. M.; Gurav, S. S.; Patil, B. M.; Srivastava, B.; Gaidhani, S. N., Combination of system biology to probe the anti-viral activity of andrographolide and its derivative against COVID-19. *RSC Adv* **2021**, *11* (9), 5065-5079.
  64. Komatsu, T. S.; Okimoto, N.; Koyama, Y. M.; Hirano, Y.; Morimoto, G.; Ohno, Y.; Taiji, M., Drug binding dynamics of the dimeric SARS-CoV-2 main protease, determined by molecular dynamics simulation. *Sci Rep* **2020**, *10* (1), 16986.
  65. Ngo, S. T.; Tam, N. M.; Pham, M. Q.; Nguyen, T. H., Benchmark of Popular Free Energy Approaches Revealing the Inhibitors Binding to SARS-CoV-2 Mpro. *J. Chem. Inf. Model.* **2021**, *61* (5), 2302-2312.
  66. Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T., Comprehensive Evaluation of Ten Docking Programs on a Diverse Set of Protein-Ligand Complexes: the Prediction Accuracy of Sampling Power and Scoring Power. *Phys Chem Chem Phys* **2016**, *18* (18), 12964-12975.
  67. Wierbowski, S. D.; Wingert, B. M.; Zheng, J.; Camacho, C. J., Cross-Docking Benchmark for Automated Pose and Ranking Prediction of Ligand Binding. *Protein Sci.* **2020**, *29* (1), 298-305.
  68. Eberhardt, J.; Santos-Martins, D.; Tillack, A.; Forli, S., AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *chemrxiv*, 2021.

