

# Novel Receptor Surface Approach for 3D-QSAR: The Weighted Probe Interaction Energy Method

Chong Hak Chae,<sup>†,‡</sup> Sung-Eun Yoo,<sup>‡</sup> and Whanchul Shin<sup>\*,†</sup>

School of Chemistry, Seoul National University, Seoul 151-742, Korea, and The Center for Biological Modulators, Korea Research Institute of Chemical Technology, 100 Jang-dong, Yuseong-gu, Daejeon 305-343, Korea

Received April 16, 2004

A 3D-QSAR technique, called the WeP (weighted probe interaction energy) method, has been developed based on the notion that certain regions of the receptor surface contribute, to varying extents, to the differences in the activities of the ligands, while other regions do not. The probes, placed around the surface of a superimposed set of ligands, were associated with fractional weights, and then an optimal distribution of probe weights that accounts for the activity profile of the training ligands was determined using a genetic algorithm. It has been shown for the three test samples that the pseudoreceptors, which consist of the surviving probes with nonzero weight values, have good predictabilities. Especially, in the case of dihydrofolate reductase inhibitors, the pseudoreceptor resembles the real protein in that there is no surviving probe in the solvent-exposed region.

## INTRODUCTION

In the field of drug design, three-dimensional (3D) quantitative structure–activity relationship (QSAR) techniques have been widely used for lead optimization when the receptor structure is not known.<sup>1</sup> From the mathematical viewpoint, the derivation of 3D-QSAR is intrinsically a difficult task since there are many unknowns but only a small quantity of observed data, and thus the problem is highly underdetermined.<sup>2</sup> There are numerous approaches for 3D-QSAR that differ in terms of field values and their sampling methods, regression methods, and variables for optimization.

CoMFA (comparative molecular field analysis),<sup>3</sup> which was first introduced a decade and a half ago, and the subsequently developed variant methods<sup>4–10</sup> have constituted the leading paradigm for 3D-QSAR techniques. In these methods, molecular fields are evaluated at regularly spaced 3D grid points around the aligned drug molecules and are correlated with the biological activities mostly by use of multivariate partial least-squares (PLS) methods.<sup>11</sup> The steric and electrostatic interaction energies between the ligands and certain types of probe are most commonly used as field values, while the hydrogen-bonding propensity<sup>12</sup> and hydrophobicity<sup>13</sup> may additionally be used. In CoMSIA (comparative molecular similarity analysis), molecular similarities, in shape, electrostatic potential, and hydrophobicity, are used as field values.<sup>14</sup> The combined utilization of CoMFA and CoMSIA has been increasing in recent years.<sup>15–17</sup> Instead of molecular fields, mutual molecular similarities between the ligands can be correlated with the activities by use of PLS (3D-QSiAR)<sup>18</sup> or genetic neural networks (SM/GNN).<sup>19,20</sup> HASL<sup>21</sup> and SOMFA<sup>22</sup> are grid-based techniques

that directly utilize the intrinsic molecular properties instead of field values and do not rely on multivariate regression methods to develop the QSAR models. In other methods, molecular descriptors, either dependent or independent of the alignment, are extracted from the internal molecular properties and correlated with the activities by use of neural networks, principal component analysis, and/or PLS.<sup>23–27</sup> Instead of molecular fields, mutual similarity for pairs of molecules can be correlated with the biological activities by use of PLS (3D-QSiAR)<sup>18</sup> or genetic neural network (SM/GNN).<sup>19,20</sup> In these methods, a molecular similarity matrix is built by comparing each compound in the data set to all the others on the basis of the values of the chosen property at 3D grid points. HASL<sup>21</sup> and SOMFA<sup>22</sup> are the grid-based techniques that utilize the intrinsic molecular properties instead of probe interaction energies and do not rely on multivariate regression methods to develop the QSAR models. Also molecular descriptors, either dependent or independent of the alignment, can be extracted from the internal molecular properties and correlated with the activities by use of neural network, principal component analysis, and/or PLS.<sup>23–27</sup>

Another approach to the 3D-QSAR studies is to utilize a hypothetical pseudoreceptor that is modeled on an isosurface for the shape field instead of rectangular 3D grids. The pseudoreceptor may be used in two different ways for the derivation of a QSAR model, exemplified by the Receptor Surface Model (RSM)<sup>28,29</sup> and the Genetically Evolved Receptor Model (GERM)<sup>30</sup> methods. In the RSM method, a receptor surface is constructed using only a set of the most active compounds by assuming complementarities of shape and properties between the receptor and the ligands. Then various 3D energy descriptors for the ligands are calculated using the pseudoreceptor as a fixed template and, together with the topological descriptors, are correlated with the activities using multivariate analysis techniques as in the conventional QSAR techniques. In the GERM method, probe

\* Corresponding author phone: +82-2-880-6656; fax: +82-2-889-5304; e-mail: nswcshin@plaza.snu.ac.kr.

<sup>†</sup> Seoul National University.

<sup>‡</sup> Korea Research Institute of Chemical Technology.

atoms of various types or properties are allowed to be situated on the surface shell of the superimposed ligands. Then the optimal distribution of the probe types, which leads to the best correlation between the calculated and observed activities, is determined by use of a genetic algorithm (GA), assuming a linear relationship between the probe interaction energies and the activities. The GERM method was later enhanced to produce the PARM (Pseudo Atomic Receptor Model) and FLARM (Flexible Atom Receptor Model) methods.<sup>31–34</sup> The Quasar method, based on the quasi-atomistic modeling concept, is also similar to the GERM in the basic framework but uses a sophisticated scheme for energy evaluation and provides complex facilities that enable the flexibilities of the receptor and ligands to be taken into account.<sup>35–37</sup> These receptor surface methods provide a comprehensive model for drug design, but the actual applications thus far have been quite limited compared to those employing CoMFA-like techniques.

In this article, we present a new receptor surface technique called the WeP (Weighted Probe interaction energy) method. It has been developed based on the notion that, upon binding of the ligands, certain regions of the receptor contribute, with varying extents, to the differences in the biological activities of the ligands, while other regions do not. For instance, the regions totally exposed to the solvent would not contribute to the activity profile. This viewpoint could be mathematically explored to derive a new 3D-QSAR algorithm by utilizing the weight-associated probes, placed around the surface of a superimposed set of ligand molecules, as a receptor surrogate. Assuming that the weighted interaction energies between the probes and the ligands correlate linearly with the activities, the optimal probe weights could be determined by minimizing the difference between the observed and calculated activities of the ligands in the training set. This optimization problem is highly underdetermined, as is typical for all 3D-QSAR paradigms, as well as being recursive due to the unknown linear regression coefficients. To solve this problem, a simple, yet practical scheme was devised that uses the GA and least-squares fitting recursively. A set of probes that survive with nonzero weight values after optimization becomes a pseudoreceptor that accounts for the activity profile of the training ligands and, at the same time, possesses a predictive ability.

The WeP method is ostensibly similar to GERM,<sup>30</sup> FLARM,<sup>33,34</sup> or Quasar<sup>35</sup> in many respects but differs in one important aspect, namely, the kind of variable employed. In the WeP method, it is the weight of each surface probe that is to be optimized, while in other methods it is the atom type or the property to be assigned to each probe. In retrospect, our method is more similar in its basic concept to the COMBINE (comparative binding energy) method which, given structural data of the ligand–receptor complexes, selects and weighs the interaction energies using PLS to deduce QSARs.<sup>38,39</sup> The WeP method has been validated and tested using three well-known samples, including the “benchmark” steroids for corticosteroid-binding globulin and a set of dihydrofolate reductase inhibitors as well as the hydrophobic chlorinated dibenzofurans, the effectors of aryl hydrocarbon receptor. Especially, in the case of the dihydrofolate reductase inhibitors, the pseudoreceptor resembles a real protein in that there is no surviving probe in the solvent-exposed region.

## METHODS

**1. Outline of the WeP Method.** In the absence of a receptor structure, a set of surface probes that are constructed to surround all of the aligned ligands can be used as the initial surrogate of the receptor binding site. Each specific region of the binding site contributes differently to the total binding energy. This feature can be represented by associating each surface probe with a weighting factor ( $w$ ) that determines the degree of contribution of that probe to the binding energy. Then, given the atomic properties of the probes, the total weighted interaction energy,  $E^{\text{WeP}}$ , between the probes and a ligand with  $n$  atoms can be defined as the following

$$E^{\text{WeP}} = \sum_j^{\text{probes}} w_j \sum_n^{\text{atoms}} E_{nj} = \sum_j^{\text{probes}} w_j E_j$$

where  $w_j$  is the weight of the probe  $j$  and  $E_j$  is the interaction energy of the probe  $j$  with the ligand. We assumed that, for the ligand molecule  $i$ , the weighted-probe interaction energy  $E_i^{\text{WeP}}$  is linearly related to its activity  $A_i$  (usually expressed as  $\text{p}K_i$ ,  $\text{pIC}_{50}$ , or  $\text{pEC}_{50}$ ).

$$A_i^{\text{cal}} = a + bE_i^{\text{WeP}} = a + b \sum_j^{\text{probes}} w_j E_{ij}$$

Then the optimal set of probe weights can be obtained by minimizing the sum of the squares of the difference ( $\chi^2$ ) between the observed and calculated activities for all ligands in the training set.

$$\chi^2 = \sum_i^{\text{ligands}} [A_i^{\text{obs}} - A_i^{\text{cal}}]^2 = \sum_i^{\text{ligands}} [A_i^{\text{obs}} - (a + b \sum_j^{\text{probes}} w_j E_{ij})]^2$$

The optimization problem becomes not only highly underdetermined due to the presence of a large number of probes but also recursive since the linear regression coefficients,  $a$  and  $b$ , are unknown a priori. We have solved this problem by using a GA and least-squares fitting in a recursive manner. In this study, the probe weight was represented as a fractional numeral in multiples of 1/15 including 0 or 1 (vide infra), to make the problem possess a combinatorial nature appropriate for the use of GA. Optimization results in a collection of surviving probes with nonzero weight values, which becomes a hypothetical pseudoreceptor that is used to predict the activities of the ligands in the test set.

A flow diagram of the WeP procedure is presented in Figure 1. The method was implemented in a program written in standard ANSI–C language. Many adjustable parameters were empirically optimized using the three test data sets. Detailed procedures are described below.

**2. Generation of the Probe Sites.** For the generation of the surface probes, we used the marching cube algorithm which is an efficient method for constructing a smooth isosurface from a 3D field of values.<sup>40,41</sup> This algorithm initially subdivides the 3D space into a series of small cubes and then leads us to “march” through each of the cubes, testing the corner points and replacing the cube with an appropriate set of triangles. The probes were generated at the vertices of the triangulated surfaces that envelop all of the aligned ligands. A probe radius of 2.0 Å, that corresponds

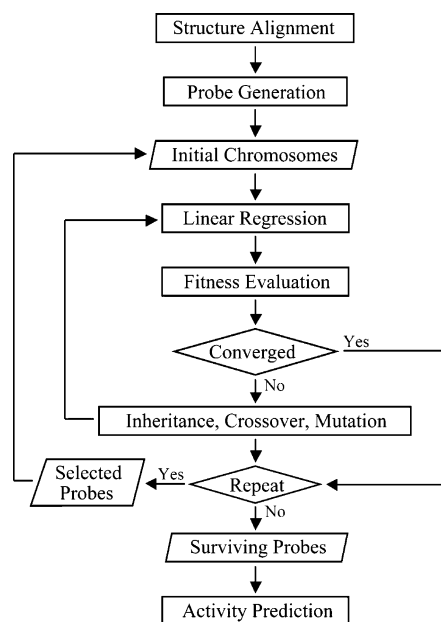


Figure 1. Flow diagram of the WeP method.

to the van der Waals radius (vdW) of a methyl group, and an initial grid size of 2.0 Å were used as default values. The average point density on the surface was relatively uniform, although the probe sites were not always spaced evenly. When two probes were situated too closely, they were merged into one at their mean position. The separations between the neighboring probes ranged from 1.8 to 2.3 Å in the final sets of surface probes for the three test cases.

### 3. Calculation of the Probe-Ligand Interaction Energy.

To evaluate the interaction energy between the probes and a ligand, the probes have to be assigned atomic properties somehow. We took an arbitrarily simplistic approach that might be adequate for the present study. All probes were given the same vdW parameters, i.e., the vdW radius ( $r_o$ ) and well depth ( $\epsilon_o$ ), of a methyl group. However, each probe was individually assigned a partial charge, which was calculated assuming that the charge distribution of the surface probes may well be complementary to the average electrostatic potential of the aligned ligands. The potential exerted by a ligand atom was calculated using the formulas  $q/r$ , where  $q$  is the partial atomic charge and  $r$  is the atom-probe distance. The potential values from all ligand atoms were summed at each probe site. These values at the probe sites were then linearly scaled to make the absolute magnitude of the maximum potential, either positive or negative, become 0.75. The scaled potential at each probe site, after its sign was reversed, was assigned as the partial charge of that probe. The resulting probe charges smoothly varied in their magnitudes over the 3D surface. The partial atomic charges of the ligand molecules were obtained using Gasteiger's PEOE (partial equalization of orbital electronegativity) method<sup>42</sup> implemented in an in-house modeling program.<sup>43</sup>

The probe-ligand interaction energies, consisting only of the vdW and charge interaction energies, were calculated using the AMBER86 force fields.<sup>44</sup> These two energies were calculated using the Lennard-Jones 12–6 potential and a distance-dependent dielectric constant, respectively. For the ligand molecules, only the non-H atoms were included in

the energy calculation utilizing the united atom force fields. This was mainly to minimize the conformational effects of the H atoms, as their positions are often ambiguous, especially in freely rotatable groups such as  $-\text{CH}_3$ ,  $-\text{NH}_3^+$ , or  $-\text{OH}$ . AMBER86 treats only the  $-\text{CH}_n$  groups as united atoms, lacking force fields for the other H-containing groups. Thus, when necessary, the  $r_o$  and  $\epsilon_o$  parameters for the united atoms were approximated by adding 0.1 Å and 0.01 kcal/mol, respectively, for each bonded H atom to the values for the parent atom. For example,  $r_o$  and  $\epsilon_o$  for the  $-\text{NH}_2$  group were estimated to be 1.95 Å and 0.18 kcal/mol, respectively, from the values of 1.75 Å and 0.16 kcal/mol for the N atom. Likewise, the partial charge of a united atom was obtained by adding those of the parent and all bonded H atoms.

**4. Optimization.** To find the optimal sets of the probe weights, the following procedures were devised. By definition, the probe weight could be any infinite decimal  $f$  in the range of  $0 \leq f \leq 1$ . We instead used fractional numbers as the weight values for computational efficiency, although the accuracy should be sacrificed somehow. In general, an  $n$ -bit binary number can denote any discrete fractional value in multiple steps of  $1/(2^n-1)$  if its decimal equivalent is divided by  $2^n-1$ . Experimentation indicated that the probe weights expressed with 4-bit bytes were most appropriate for our purpose considering the overall accuracy and computational efficiency. Accordingly, we allowed each probe to take a weight value out of the sixteen fractional numbers of 0, 1/15, 2/15, ..., 14/15, and 1. With this binary representation, a weight value can easily be changed by a bit operation, which is quite suitable for GA optimization. Consequently, a chromosome used in this study is a binary string that consists of a series of the 4-bit genes, each of which encodes the weight value of a specific surface probe.

The fitness function for the phenotype  $k$  was defined as  $\chi(k)^2$ .

$$\chi(k)^2 = \sum_i^{\text{ligands}} [A_i^{\text{obs}} - A(k)_i^{\text{cal}}]^2 = \sum_i^{\text{ligands}} [A_i^{\text{obs}} - (a(k) + b(k)E_i^{\text{WeP}})]^2 = \sum_i^{\text{ligands}} [A_i^{\text{obs}} - (a(k) + b(k) \sum_j^{\text{probes}} w(k)_j E_{ij})]^2$$

Thus the smaller the  $\chi^2$  value, the better the chromosome. Given a set of the probe weight  $w_j$ , the linear regression coefficients  $a$  and  $b$  can be determined by the least-squares method fitting the  $A_i^{\text{obs}}$  data and the calculated  $E^{\text{WeP}}$  values for the ligands in the training set. The activity data used in this study were expressed as  $-\log(K_i)$ ,  $\text{IC}_{50}$ , or  $\text{EC}_{50}$ . Therefore, the calculated activity takes on higher values as the  $E^{\text{WeP}}$  decreases, implying that the slope  $b$  should take a negative value for the correct solutions. The regression coefficients have to be reevaluated for each chromosome whenever the  $w_j$ 's are changed during evolution. It is noteworthy that the interaction energy  $E_{ij}$  between the ligand  $i$  and the probe  $j$  needs to be evaluated only once at the onset of the whole procedure.

The optimization of  $\chi^2$  was performed using GA.<sup>45</sup> Initial populations were made by randomly assigning 0 or 1 to each bit of the chromosomes. The optimal population size usually



depends on the length of the chromosome, i.e., the number of the surface probes, and one or two thousand chromosomes were used in this study (see the Results). For each chromosome, the  $E^{\text{WeP}}$  value, the regression coefficients  $a$  and  $b$ , and the  $\chi^2$  value were determined in that sequence. Then the chromosomes were sorted with respect to  $\chi^2$ . The chromosomes within the smallest 5% of the  $\chi^2$  values were selected as the "elitists" and passed on to the next generation. If a chromosome had a positive slope  $b$ , it was not selected as an elitist. The remaining members, amounting to 95% of the population size, were newly generated by crossover using the tournament selection and uniform crossover schemes. Finally 1% of the total chromosomes were randomly selected and mutated by flipping a randomly selected bit. During evolution, a newly made individual was always checked against the existing pool of chromosomes and accepted only if there was no duplicate. This check for nonredundancy was to enforce genetic diversity at each generation and thus to expedite a wider exploration of the combinatorial hyperspace. For a new generation, then, the processes of evaluation of  $E^{\text{WeP}}$ , least-squares fitting, and GA optimization were performed again.

These optimization cycles were repeated until the convergence criteria were met. In this study, the system was considered to be converged when the  $\chi^2$  values of the elitists ceased to decrease further and, at the same time, the *rms* differences of the weight values between two successive generations consistently remained smaller than 0.001 for the preceding 100 generations. For all three test cases, the first rounds of optimization cycles were finished before 1000 generations, initially specified as the maximum number of cycles. No assumption was made nor any manipulation carried out in the determination of the regression coefficients  $a$  and  $b$ . They were determined solely by the weight values via  $E^{\text{WeP}}$  for every chromosome generated during evolution. These values varied widely among the chromosomes at the initial stage of optimization but rapidly approached the correct ones as the system evolved.

Upon convergence, the elitists showed very similar distributions of probe weights with only small variations. They also had very similar regression coefficients. However, we did not take a single elitist as the final solution. Instead, as is usually done in optimization studies using GA, we made an averaged model from a collection of the elitists, to eliminate the possibility of picking up a spurious or overfitted solution that might accidentally appear among the elitists. In a highly underdetermined system of linear equations, there is always a possibility that the best-scored solution is actually an overfitted one without any predictive power. The statistically averaged weight of the probe  $j$  was obtained from the following formulas using only the elitists

$$w_j^{\text{av}} = c \sum_k^{\text{elitists}} w_{jk} \times S_k = c \sum_k^{\text{elitists}} w_{jk} \times \exp[-(\chi_k^2 - \chi_{\min}^2)/RT]$$

where  $S_k$  represents the relative weight of the chromosome  $k$  and  $c$  is a scaling constant to make the maximum average weight become 1.0. The  $S_k$  values were estimated from the Boltzmann distribution, in which  $\chi_{\min}^2$  is the value for the best chromosome. The constant  $RT$  was arbitrarily set as 0.4343, with which  $S_k$  becomes 0.5 when  $\chi_k^2 - \chi_{\min}^2$  equals 0.1.

After the first round of optimization was finished, surface probes with significant weight values were selected, and another round of optimization was performed using only these probes. In this study, the probes with an average weight larger than 0.01 were considered significant. For this reduced subset of the surface probes, the optimizations were performed in exactly the same way as the first round, starting with new chromosomes made with completely randomized initial weights. This whole process, as was empirically found, resulted in a better fitness, albeit slightly, and converged very quickly.

All surface probes that had nonzero average weight values upon convergence were labeled as the "surviving" probes. A collection of the surviving probes became a hypothetical pseudoreceptor, which can be used for the activity prediction. The overall regression coefficients were obtained by averaging the values for all elitists. The final WeP results consisted of the coordinates, a partial charge, and the average weight value for each of the surviving probes and the coefficients  $a$  and  $b$ . As is usually done in other 3D-QSAR studies, the quality of the training was judged by the following squared correlation coefficient  $r^2$ , where  $A^{\text{mean}}$  is the mean value of observed activities.<sup>46</sup>

$$r^2 = 1 - \frac{\sum (A_i^{\text{obs}} - A_i^{\text{cal}})^2}{\sum (A_i^{\text{obs}} - A^{\text{mean}})^2}$$

**5. Prediction.** The predictive power of the pseudoreceptor was examined using the test ligand molecules, with known activities, that were superimposed onto the training molecules but were not included in the training. The weighted probe interaction energy ( $E^{\text{WeP}}$ ) between each test ligand and the surviving probes was calculated and then linearly transformed into the activity ( $A^{\text{pred}}$ ) using the overall regression coefficients  $a$  and  $b$ . If a test ligand was significantly larger than the training molecules, it would probably collide with the pseudoreceptor, yielding an unrealistically high interaction energy. Therefore, we took precautionary measures so that the activity prediction could provisionally be made ignoring the interaction energies higher than 0.0 kcal/mol. Detailed information on close contacts, if present, is given so that the conformation or alignment of the test ligand can be adjusted to avoid deleterious contacts. In the three samples studied here, however, no test ligand made collisions with the pseudoreceptors.

The quality of prediction for a set of test ligands was judged by the predictive correlation coefficient  $r_{\text{pred}}^2$ , which has the same formulas as  $r^2$  except in that  $A^{\text{cal}}$  is replaced by  $A^{\text{pred}}$ . The denominator in  $r_{\text{pred}}^2$  is the sum of the squared prediction errors, which is called the *PRESS* (predictive residual sum of squares). The standard deviation of errors of prediction (*SDEP*) was calculated from  $\text{SDEP} = \text{SQRT}(\text{PRESS}/n)$  where  $n$  is the number of molecules in the test set.

## RESULTS

The WeP method has been tested with three well-known samples. The first sample is a highly congeneric series of chlorinated dibenzofurans that function as effectors of aryl hydrocarbon receptor (AHR).<sup>47</sup> The second is a set of steroid compounds that exhibit a range of binding affinities for

corticosteroid-binding globulin (CBG). This data set has become a benchmark for testing 3D-QSAR methods since the first CoMFA study of Cramer et al.<sup>3,48</sup> The third sample is a series of dihydrofolate reductase (DHFR) inhibitors that are relatively diverse in their structures.<sup>21</sup> Generation of the atomic coordinates and alignment of the molecules for the AHR and DHFR data sets were performed using an in-house modeling program.<sup>43</sup>

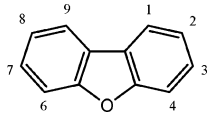
During the development of a 3D-QSAR method, one has to ascertain that the molecules in the training set are well distributed over the structural spaces, not strongly clustered, and that the molecules in the prediction set are not very close to the training molecules. To examine how widely the WeP results vary depending on the selection of the training molecules and also to ensure that the results are free from chance correlation or overfitting, we extensively performed the permutation and scramble tests for each of the three samples. In the permutation tests, the trainings and predictions were performed for 10 different sets of the molecules with original activity data. One of the training sets, called the representative set, contained the molecules that were somewhat judiciously selected, and the remaining sets contained the randomly selected training molecules. The scramble (or y-randomization) tests were performed 10 times for every training set used for the permutation tests. For each of the data sets used for the scramble tests, the original activity data were randomly exchanged among the ligand molecules. For each test sample, therefore, the WeP analyses were performed 110 times in total, which may be sufficient to validate the results from a statistical point of view.

All the calculations described herein were carried out using a Linux PC with a Pentium IV 1.6 GHz CPU. It took approximately 1, 6, and 7 min, respectively, for each run of the AHR, CBG, and DHFR data sets. The execution times might be reduced if the Linux system as well as the WeP program were fully optimized. The results of the permutation and scramble tests as well as the detailed results for the representative sets are described below for the three samples.

**1. The AHR Set.** The AHR data set of 34 chlorinated dibenzofurans, taken from the article,<sup>24</sup> is listed in Table 1. The atomic coordinates were made based on the crystal structures found in the Cambridge Structural Database (CSD). This data set containing compounds that are structurally congeneric and highly hydrophobic may be too simple a case for a 3D-QSAR problem. These molecules have no conformational degree of freedom and only two different types of atom (excluding H atoms). This simplistic nature, however, was quite beneficial in initially validating whether the concept of the weighted probe interaction energy can be substantiated as a working QSAR method and also to find the optimal values for many adjustable parameters involved in the WeP method. The only problem with this data set was how to align the molecules with a fused ring system that possess mirror or 2-fold rotational symmetry. There are two alternative ways to superimpose the asymmetrically chlorinated dibenzofurans onto each other, and thus 2<sup>N</sup> superposition modes are possible for N such molecules.

To circumvent the difficulties in the alignment process, we made an initial training with six molecules that posed little ambiguity in superposition, anticipating that the predictions with the resulting pseudoreceptor might help differenti-

**Table 1.** Structures and Binding Affinities of the AHR Set

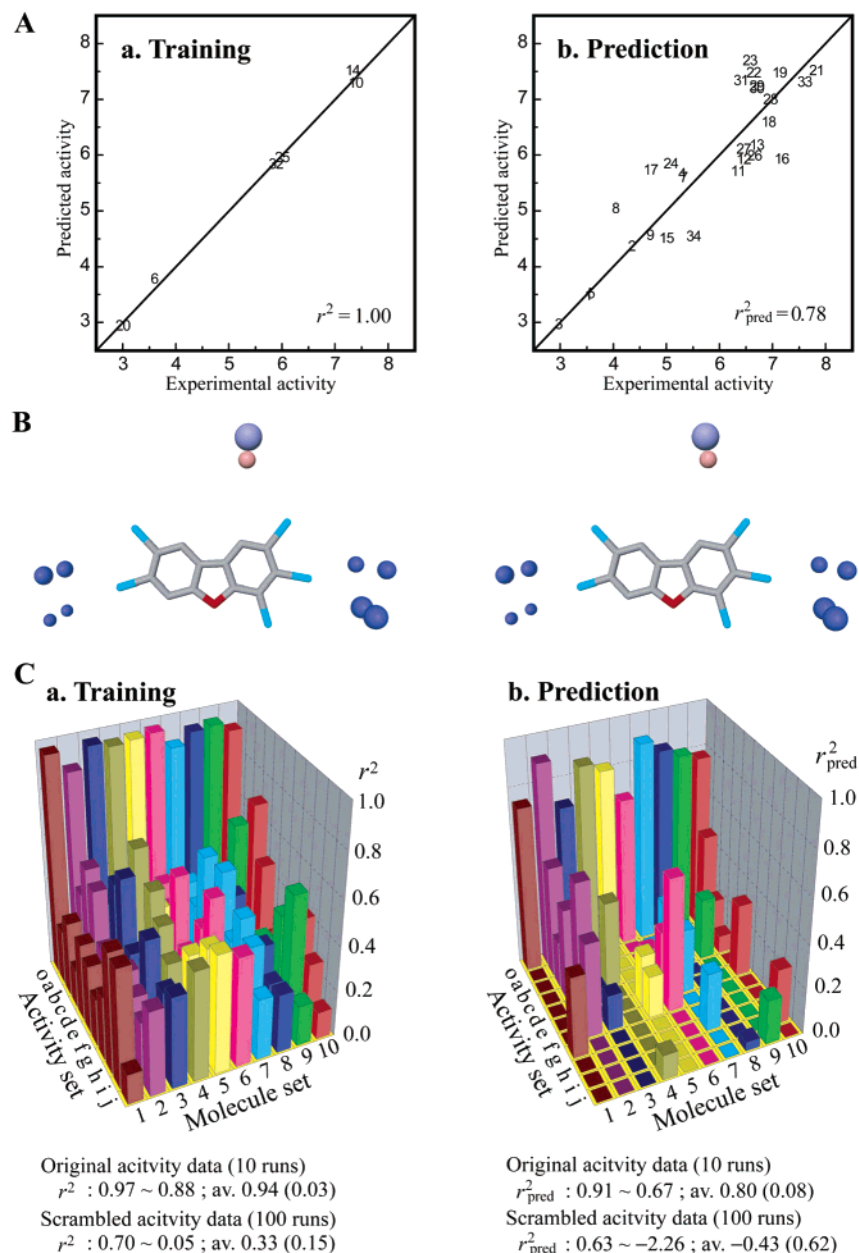


no. <sup>a</sup>	compound	$A_{obs}^b$	$\Delta A^c$	$\Delta A_r^d$
1	2-Cl	3.553	-0.089	-0.343
2	3-Cl	4.377	0.792	-0.037
3	4-Cl	3.000	0.237	-0.052
4	2,3-Cl <sub>2</sub>	5.326	0.318	-0.784
5	2,6-Cl <sub>2</sub>	3.609	-0.114	-0.080
6*	2,8-Cl <sub>2</sub>	3.590		
7	1,3,6-Cl <sub>3</sub>	5.357	0.222	-0.566
8	1,3,8-Cl <sub>3</sub>	4.071	1.775	0.951
9	2,3,4-Cl <sub>3</sub>	4.721	1.144	-0.178
10*	2,3,8-Cl <sub>3</sub>	6.000		
11	2,6,7-Cl <sub>3</sub>	6.347	-1.453	-0.656
12	2,3,4,6-Cl <sub>4</sub>	6.456	-0.558	-1.592
13	2,3,4,8-Cl <sub>4</sub>	6.699	-0.541	-1.609
14*	2,3,7,8-Cl <sub>4</sub>	7.387		
15	1,2,4,8-Cl <sub>4</sub>	5.000	-0.512	-1.024
16	1,2,4,6,7-Cl <sub>5</sub>	7.169	-1.549	-1.262
17	1,2,4,7,9-Cl <sub>5</sub>	4.699	1.010	1.331
18	1,2,3,4,8-Cl <sub>5</sub>	6.921	-0.355	-1.649
19	1,2,3,7,8-Cl <sub>5</sub>	7.128	0.551	0.324
20*	1,2,4,7,8-Cl <sub>5</sub>	5.886		
21	2,3,4,7,8-Cl <sub>5</sub>	7.824	-0.334	-0.552
22	1,2,3,4,7,8-Cl <sub>6</sub>	6.638	1.259	0.816
23	1,2,3,6,7,8-Cl <sub>6</sub>	6.569	1.112	1.102
24	1,2,4,6,7,8-Cl <sub>6</sub>	5.081	0.740	1.296
25*	2,3,4,6,7,8-Cl <sub>6</sub>	7.328		
26	2,3,6,8-Cl <sub>4</sub>	6.658	-0.691	-1.253
27	1,2,3,6-Cl <sub>4</sub>	6.456	-0.374	-1.417
28	1,2,3,7-Cl <sub>4</sub>	6.959	0.513	0.016
29	1,3,4,7,8-Cl <sub>5</sub>	6.699	0.706	0.513
30	2,3,4,7,9-Cl <sub>5</sub>	6.699	0.723	0.479
31	1,2,3,7,9-Cl <sub>5</sub>	6.398	1.167	0.914
32*		3.000		
33	2,3,4,7-Cl <sub>4</sub>	7.602	-0.314	-0.804
34	1,2,4,6,8-Cl <sub>5</sub>	5.509	-0.991	-1.216

<sup>a</sup> Compounds with an asterisk are used as the initial training set.

<sup>b</sup>  $A_{obs}$  is the experimental activity in pEC<sub>50</sub>. <sup>c</sup>  $\Delta A = A_{cal} - A_{obs}$ . <sup>d</sup>  $\Delta A_r$  for the reversed orientation; see text.

ate two alternate orientations for each of the asymmetric compounds. The training set consisted of all four symmetric molecules (**6**, **14**, **25**, **32**) in the data set and two asymmetric ones (**10**, **20**) added as a symmetry breaker. There were 107 surface probes surrounding the training molecules. GA optimizations were performed with a population size of 1000 and converged with a correlation coefficient  $r^2$  of 1.00 (Figure 2A). As shown in Figure 2B, the pseudoreceptor consists of 10 surviving probes whose partial charges and weights are listed in Table 2. To our own surprise, the WeP model trained with only six molecules had a relatively good predictability. Differences between the observed and predicted activities for 28 test compounds, each in two opposing configurations, are listed in Table 1. For each compound, one configuration showed a smaller difference than the other. A collection of these configurations was assumed, albeit somewhat arbitrarily, as the correctly aligned set of molecules. For this AHR set,  $r_{pred}^2$  was 0.78, the maximum difference being 1.262 in log units for compound **16** (Figure 2A). For comparison,  $r_{pred}^2$  for all 56 configurations was 0.50. This preliminary result encouragingly demonstrated that the present WeP procedures may function as a 3D-QSAR method.



**Figure 2.** (A) Experimental versus predicted binding affinities for (a) training and (b) prediction with correctly oriented configurations of the AHR set. (B) Stereoview of the pseudoreceptor for the AHR set obtained with six training molecules. The spheres denote the surviving probes and the most active compound **21** (hydrogen atoms omitted for clarity) was overlaid as a frame of reference. The radius ( $R$  in Å) of a sphere was linearly scaled to the probe weight ( $w$ ) according to  $R = 0.5w + 0.2$ . The probes with a positive partial charge were tinted blue and those with a negative charge red. The degree of saturation of color is approximately proportional to the absolute magnitude of the charge, the neutral probes being colored white. (C) Plots of the correlation coefficients with other statistical values for (a) training ( $r^2$ ) and (b) prediction ( $r^2_{pred}$ ) obtained from the permutation and scramble tests. Activity set o is for the permutation tests (molecule sets 1 to 10) with correct activity data, and other activity sets (a to j) are for the scramble tests. Negative  $r^2$  and  $r^2_{pred}$  values were ignored and not shown in the figure.

With the 34 correctly aligned molecules, the permutation tests with the original activity data were then performed to check the overall stability of the WeP method. Ten different training sets, each containing 16 molecules randomly selected, were prepared. All data sets were well trained. The  $r^2$  values ranged from 0.88 to 0.97 with an average value of 0.94 and standard deviation of 0.03. When the activities for the remaining 18 test molecules were predicted for each data set,  $r^2_{pred}$  ranged from 0.67 to 0.91 with an average value of 0.80 ( $\pm 0.08$ ) without any significant outlier. The  $r^2$  and  $r^2_{pred}$  values from permutation tests are plotted in Figure 2C

(activity set o). The numbers and locations of the surviving probe sites did not vary significantly among the test sets (data not shown). In particular, most of the heavily weighted probes commonly appeared in all test sets. These indicate that the training of the AHR data set was performed in a stable manner, not strongly depending on the selection of the training molecules, and that all WeP models possessed relatively good predictabilities.

For each training set used in the permutation tests, we prepared 10 data sets for the scramble tests by randomly exchanging the original activities of 34 molecules and



**Table 2.** Partial Charges and Weight Values of the Surviving Probes in the Pseudoreceptors for the Representative Test Sets

no.	AHR		CBG		DHFR	
	charge	weight	charge	weight	charge	weight
1	0.027	1.000	0.088	1.000	-0.375	1.000
2	0.180	0.828	0.157	1.000	-0.329	1.000
3	0.183	0.827	-0.010	1.000	-0.116	1.000
4	-0.001	0.587	0.074	1.000	0.118	1.000
5	0.204	0.572	-0.048	1.000	0.366	1.000
6	0.228	0.571	0.024	0.999	0.037	1.000
7	0.226	0.568	-0.101	0.997	0.196	0.999
8	0.199	0.539	-0.062	0.991	-0.095	0.963
9	0.180	0.268	0.088	0.989	-0.043	0.958
10	0.182	0.267	-0.210	0.985	-0.056	0.845
11			-0.116	0.969	-0.101	0.800
12			-0.156	0.934	-0.568	0.731
13			-0.135	0.880	0.036	0.683
14			0.133	0.812	0.276	0.533
15			0.139	0.800	0.133	0.533
16			-0.134	0.766	-0.312	0.533
17			0.064	0.737	0.356	0.467
18			0.179	0.600	0.141	0.400
19			-0.125	0.533	0.095	0.400
20			-0.078	0.533	-0.009	0.366
21			0.436	0.533	0.096	0.333
22			-0.162	0.533	0.566	0.272
23			-0.119	0.482	0.403	0.267
24			-0.089	0.411	0.183	0.200
25			0.050	0.271	0.042	0.068
26			-0.145	0.267	0.000	0.066
27			0.318	0.133	0.012	0.013

performed the WeP analyses 100 times in total. The  $r^2$  values for the training in each case with 16 molecules ranged from 0.05 to 0.70 with an average value of 0.33 ( $\pm 0.16$ ). These values are significantly lower than those from the permutation tests using the original activity data, indicating that all of the models were poorly trained. The results for predictions were much worse. Only 18 data sets out of 100 showed positive  $r^2_{pred}$  values, the largest being 0.60. The  $r^2$  and  $r^2_{pred}$  values from scramble tests are plotted in Figure 2C (activity sets a to j). Notably, not a single data set showed meaningful  $r^2$  and  $r^2_{pred}$  values simultaneously. These results indicate that, for the AHR set, the WeP analyses with normal data provided reliable QSARs, not suffering from chance correlations.

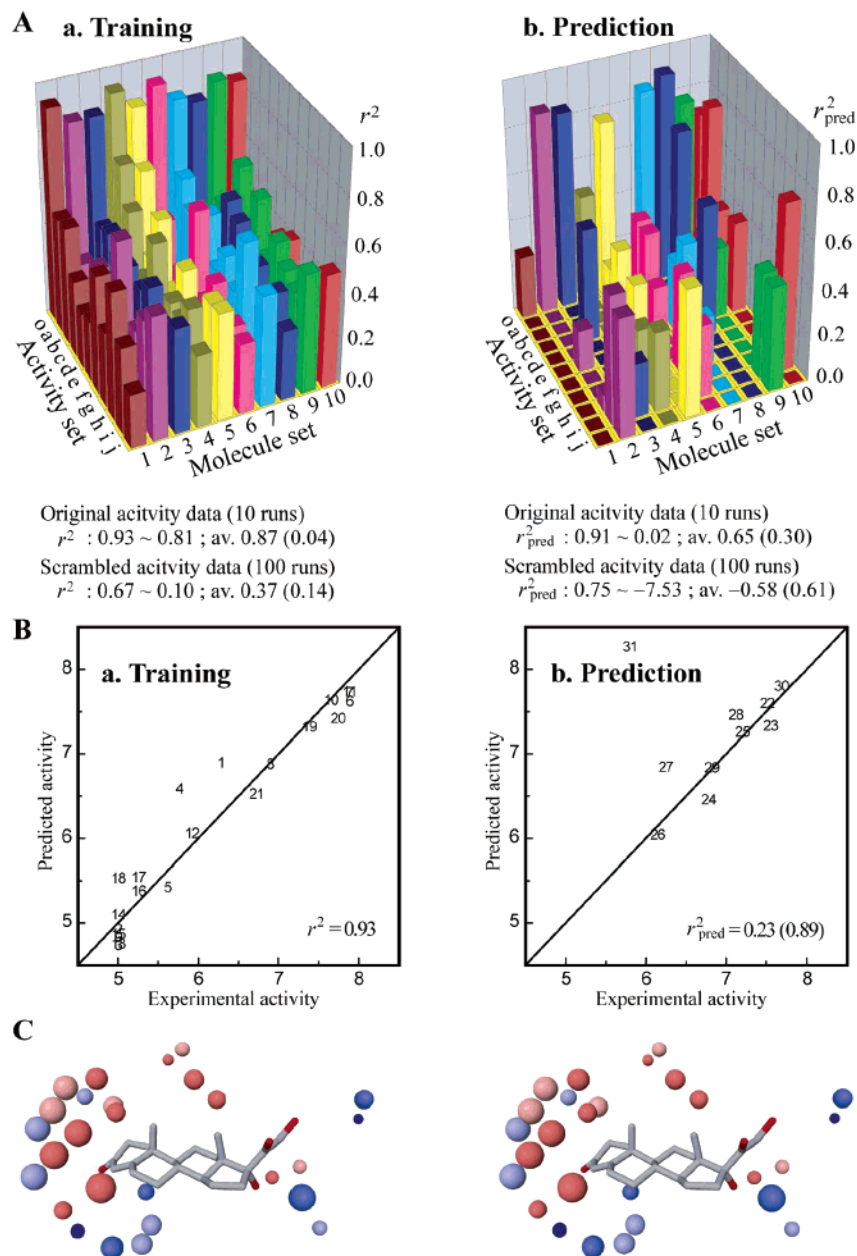
**2. The CBG Set.** The atomic coordinates of 31 steroid molecules, listed in Table 3, were downloaded from Gasteiger's Internet homepage.<sup>49</sup> Some mistakes present in the earlier studies were fixed in this data set.<sup>24</sup> Each of the molecules was aligned by the rigid-body, least-squares fitting of all 17 carbon atoms in the steroid ring to the corresponding atoms of deoxycortisol (**11**). Permutation and scramble tests were performed in the same manner as in the AHR case. Ten different training sets were prepared. Among these, the representative set (molecule set 1 in Figure 3A) contained steroids **1–21** as the training molecules and steroids **22–31** as the test molecules, in line with most published CBG studies, and other sets contained 21 training and 10 test molecules randomly selected. There were 124 surface probes surrounding the training molecules in the representative set. The number of probes varied slightly among the molecule sets, but GA optimizations were performed with an equal population size of 2000. The  $r^2$  and  $r^2_{pred}$  values from permutation (activity set o) and scramble (activity sets a to j) tests are plotted in Figure 3A.

**Table 3.** Compound Names and Binding Affinities of the CBG Set

no. <sup>a</sup>	compound	$A_{obs}$ <sup>b</sup>	$\Delta A$ <sup>c</sup>
1*	aldosterone	6.279	
2*	androstenediol	5	
3*	androstenediol	5	
4*	androstenedione	5.763	
5*	androsterone	5.613	
6*	corticosterone	7.881	
7*	cortisol	7.881	
8*	cortisone	6.892	
9*	dehydroepiandrosterone	5	
10*	deoxycorticosterone	7.653	
11*	deoxycortisol	7.881	
12*	dihydrotestosterone	5.919	
13*	estradiol	5	
14*	estriol	5	
15*	estrone	5	
16*	etiocholanolone	5.255	
17*	pregnenolone	5.255	
18*	17-hydroxypregnenolone	5	
19*	progesterone	7.380	
20*	17-hydroxyprogesterone	7.740	
21*	testosterone	6.724	
22	prednisolone	7.512	0.066
23	cortisol-21-acetate	7.553	-0.237
24	4-pregnene-3,11,20-trione	6.779	-0.338
25	epicorticosterone	7.200	0.042
26	19-nortestosterone	6.144	-0.120
27	16 $\alpha$ ,17-dihydroxy-4-pregnene-3,20-dione	6.247	0.576
28	16 $\alpha$ -methyl-4-pregnene-3,20-dione	7.120	0.323
29	19-norprogesterone	6.817	0.001
30	11 $\beta$ ,17,21-trihydroxy-2 $\alpha$ -methyl-4-pregnene-3,20-dione	7.688	0.095
31	11 $\beta$ ,17,21-trihydroxy-2 $\alpha$ -methyl-9 $\alpha$ -fluoro-4-pregnene-3,20-dione	5.797	2.455

<sup>a</sup> Compounds with an asterisk are used as the representative training set. <sup>b</sup>  $A_{obs}$  is the experimental activity in  $pK_i$ . <sup>c</sup>  $\Delta A = A_{cal} - A_{obs}$ , where  $A_{cal}$  is the activity predicted by the WeP method.

In the permutation tests with the original activity (or binding affinity) data, all data sets were decently trained independently of the training molecules selected. The  $r^2$  values ranged from 0.81 to 0.93 with an average value of 0.87 ( $\pm 0.04$ ). Predictions of activity for 10 test molecules varied widely in quality depending on the molecular sets. The  $r^2_{pred}$  values ranged from 0.02 to 0.91 with an average value of 0.65 ( $\pm 0.30$ ). Predictions for seven molecule sets gave decent results with the  $r^2_{pred}$  values greater than 0.70, but the  $r^2_{pred}$  values for three sets were quite low owing to the presence of outliers. Molecule sets 1 and 4 had  $r^2_{pred}$  values of 0.28 and 0.47, respectively, both containing steroid **31** as an outlier. Molecule set 6 had a particularly low  $r^2_{pred}$  of 0.02, and steroid **1** was a distinct outlier in this set. If these outliers were removed, however, the  $r^2_{pred}$  values became 0.93, 0.92, and 0.50, respectively, for the molecule sets 1, 4, and 6. The appearance of an outlier is a phenomenon frequently occurring in 3D-QSAR studies of the CBG set as discussed below. The scramble tests were performed 100 times as in the AHR set, and all of the models were poorly trained. The  $r^2$  values ranged from 0.10 to 0.67 with an average value of 0.37 ( $\pm 0.13$ ), which values are significantly lower than those from the permutation tests. Thirty-two molecule sets showed positive  $r^2_{pred}$  values. Several sets had relatively large  $r^2_{pred}$  values above 0.7, which might have resulted from chance correlation, but all of them were trained very poorly with  $r^2$  below 0.5.



**Figure 3.** (A) Plots of the correlation coefficients with other statistical values for (a) training ( $r^2$ ) and (b) prediction ( $r^2_{pred}$ ) obtained from the permutation and scramble tests. See the legends in Figure 2(C) for the detailed descriptions of the plots. (B) Experimental versus predicted binding affinities for (a) training and (b) prediction of the representative CBG set.  $r^2_{pred}$  in parentheses is the value when steroid **31** was excluded. (C) Stereoview of the pseudoreceptor for the representative CBG set. Steroid **11** was overlaid as a frame of reference. See the legends in Figure 2(B) for the detailed descriptions of the spheres.

The results for the representative set (molecule set 1) deserve a detailed discussion compared with those from other 3D-QSAR studies. The WeP model was trained with an  $r^2$  of 0.93 (Figure 3B). The leave-one-out cross-validated correlation coefficient ( $q^2$ ) was 0.76. The pseudoreceptor for the representative set is shown in Figure 3C. There are 27 surviving probes whose charge values and weights are listed in Table 2. Most of the surviving probes with large weight values have nearly neutral charges. Differences between the observed and predicted activities for steroids **22**–**31** obtained from various studies are compared in Table 4. In the WeP prediction, the fluorinated steroid **31** behaved as a significant outlier as in other predictions. Its activity was overestimated by far (2.5 log units). If this compound was removed, the  $r^2_{pred}$  and SDEP values became 0.93 and 0.27 from 0.28 and

0.82, respectively. These values are comparable to or even better than those from other 3D-QSAR methods, which demonstrates the predictive ability of the WeP method.

Steroid **31** is a notorious anomaly in activity prediction using 3D-QSAR methods. A reason for this behavior may be sought by comparing the structures and activity data of closely related compounds. Steroid **31** is most similar in structure to steroid **30**, the only difference being the F and H atoms at the 9 carbon, respectively, and then to steroid **7** in the training set. The observed activities of steroids **7** and **30** are similar with values of 7.881 and 7.688, respectively. The activity of steroid **30** was predicted to be 7.783, showing a good correlation between structure and activity. The observed activity of steroid **31** is 5.797 and is smaller by about 2 log units. This means that the binding energy of



**Table 4.** Comparison of the WeP Results with Other Highly Predictive 3D-QSAR Techniques for the CBG Set<sup>a</sup>

no.	$A_{obs}$	WeP <sup>b</sup>	CoMFA <sup>a</sup>	CoMFA (FFD) <sup>a</sup>	SM analysis <sup>a</sup>	Compass <sup>c</sup>	MS-WHIM <sup>d</sup>	SOMFA <sup>a</sup>	GRIND <sup>e</sup>
22	7.512	0.066	0.572	0.371	-0.059	-0.450	-0.212	-0.233	0.368
23	7.553	-0.237	0.113	-0.123	-0.531	0.176	0.779	-0.519	0.176
24	6.779	-0.338	-0.241	-0.137	0.160	-0.317	0.042	0.146	0.094
25	7.2	0.042	0.604	0.505	-0.054	0.266	0.245	0.032	0.357
26	6.144	-0.120	0.252	0.351	-0.236	-0.150	-0.023	-0.400	-0.438
27	6.247	0.576	1.099	0.715	0.799	0.136	0.654	0.553	0.258
28	7.12	0.323	-0.110	-0.272	-0.551	-0.495	-0.588	-0.517	-0.070
29	6.817	0.001	0.047	-0.001	0.033	0.586	0.021	-0.125	0.216
30	7.688	0.095	0.282	0.079	-0.149	0.053	0.172	-0.343	-0.082
31	5.797	2.455	2.208	1.996	1.660	1.982	1.694	1.486	
$r^2_{pred}$		0.278	0.241	0.445	0.557	0.462	0.525	0.629	
(0.930)		(0.769)	(0.876)	(0.855)	(0.888)	(0.836)	(0.869)	(0.933)	(0.769)
SDEP		0.816	0.837	0.716	0.640	0.705	0.662	0.584	
(0.266)		(0.486)	(0.356)	(0.385)	(0.339)	(0.411)	(0.367)	(0.262)	(0.486)

<sup>a</sup> This table was made based in Table 4 of ref 22. The values of  $\Delta A$  ( $= A_{cal} - A_{obs}$ ) are listed instead of the predicted activities ( $A_{cal}$ ). In parentheses are the values of  $r^2_{pred}$  and SDEP when steroid **31** is excluded. <sup>b</sup> Present study. <sup>c</sup> Reference 23. <sup>d</sup> Reference 25. <sup>e</sup> Reference 26.

steroid **31** would be higher by about 2.6 kcal/mol than the other two steroids. It is doubtful whether a single F atom at the 9 position can cause such a detrimental effect in the binding of a steroid to the CBG protein. Considering the structural differences, therefore, it is more likely that the activities of steroids **30** and **31** would differ marginally as predicted in most QSAR studies. A better reason for the anomaly of steroid **31** has already been alluded to in the different experimental techniques used to calculate the CBG binding affinities.<sup>22</sup> In the present study, the average  $r^2$  of the CBG set is lower (0.87 vs 0.94) in the permutation tests and higher (0.37 vs 0.33) in the scramble tests than those of the AHR set. These observations may reflect the inconsistency possibly present within the experimental binding affinities in the CBG data set.

**3. The DHFR Set.** The DHFR set of 72 inhibitors that encompass a variety of structures is listed in Table 5. This data set was taken from the article of HASL which contains data compiled from several sources.<sup>21</sup> The atomic coordinates were made based on the conformation of folic acid observed in the human DHFR-folate complex (PDB code, 1dhf),<sup>50</sup> and the energy of each molecule was minimized. The molecules were aligned by superimposing heterocyclic rings. The permutation and scramble tests were performed with 10 different training sets. In the representative set (molecule set 1), 36 odd-numbered compounds were used as the training molecules and the remaining ones as the test molecules. The training molecules were more or less evenly distributed in terms of the properties and sizes of the substituents. Other molecule sets contained 36 training and 36 test molecules randomly selected. The number of probes ranged from 225 to 230 depending on the molecule sets, but GA optimizations were performed with an equal population size of 2000.

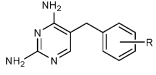
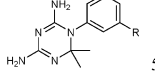
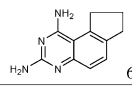
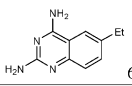
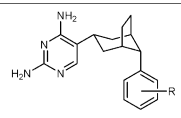
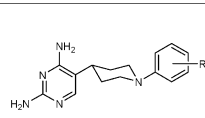
The  $r^2$  and  $r^2_{pred}$  values from permutation (activity set o) and scramble (activity sets a to j) tests are plotted in Figure 4A. In the permutation tests with the original activity, the  $r^2$  values ranged from 0.75 to 0.92. The average  $r^2$  was 0.83 ( $\pm 0.05$ ) which is lower than the values 0.94 ( $\pm 0.03$ ) and 0.87 ( $\pm 0.04$ ) for the AHR and CBG sets, respectively. The results for predictions were worse than those for the other two data sets. The average  $r^2_{pred}$  was 0.51 ( $\pm 0.14$ ), which is significantly lower than the values 0.80 and 0.65 obtained for the AHR and CBG sets, respectively. The maximum

$r^2_{pred}$  was 0.70 which is slightly lower than the value of 0.75 obtained with HASL.<sup>21</sup> This rather poor performance for the DHFR case might be due to the structural diversity present in its data set. In each prediction set, two or three out of 36 compounds behaved as the significant outliers showing the maximum difference of 3 log units. When these outliers were removed, the average  $r^2_{pred}$  became 0.58 ( $\pm 0.09$ ). The scramble tests indicated that the training (average  $r^2$ , 0.35 ( $\pm 0.14$ )) could be performed to a degree similar to that for the AHR and CBG sets, but all predictions were virtually meaningless as shown in Figure 4A(b). The maximum  $r^2_{pred}$  was 0.38. These results clearly indicate that the WeP analyses of the DHFR set with the original activity data were not prone to chance correlation.

For the representative set, the WeP model converged with an  $r^2$  of 0.80 as shown in Figure 4B(a). The leave-one-out cross-validated correlation coefficient  $q^2$  was 0.53. The pseudoreceptor, shown in Figure 4C, consisted of 32 probes surviving out of the initial 228 probe sites. The partial charges and weights of the surviving probes are listed in Table 2. Contrary to the CBG and AHR cases, the heavily weighted probes generally had large partial charges, either positive or negative, suggesting that the electrostatic interactions involving the charged or polar groups might be the most important factors in determining the activity differences for the DHFR ligands. The activity prediction for the 36 even-numbered test compounds, shown in Figure 4B(b), gave an  $r^2_{pred}$  value of 0.57 and an SDEP of 1.03. When the compounds **54** and **66**, two distinct outliers showing a difference of 3 log units in the activity, were removed, the  $r^2_{pred}$  and SDEP values became 0.78 and 0.67, respectively. It was encouraging that the DHFR data set containing somewhat dissimilar classes of molecules yielded a relatively good prediction with an  $r^2_{pred}$  value comparable to those obtained for other data sets containing the molecules relatively congeneric in their shapes and overall sizes.

The pseudoreceptors obtained from the permutation tests are similar to each other. In Figure 5, all of the surviving probes from the 10 data sets and all of the aligned ligands were superimposed onto the crystal structure of the DHFR-folate complex.<sup>50</sup> This figure clearly shows that the surviving probes form many clusters that are scattered around the binding cleft of the protein. Moreover, there is a wide region

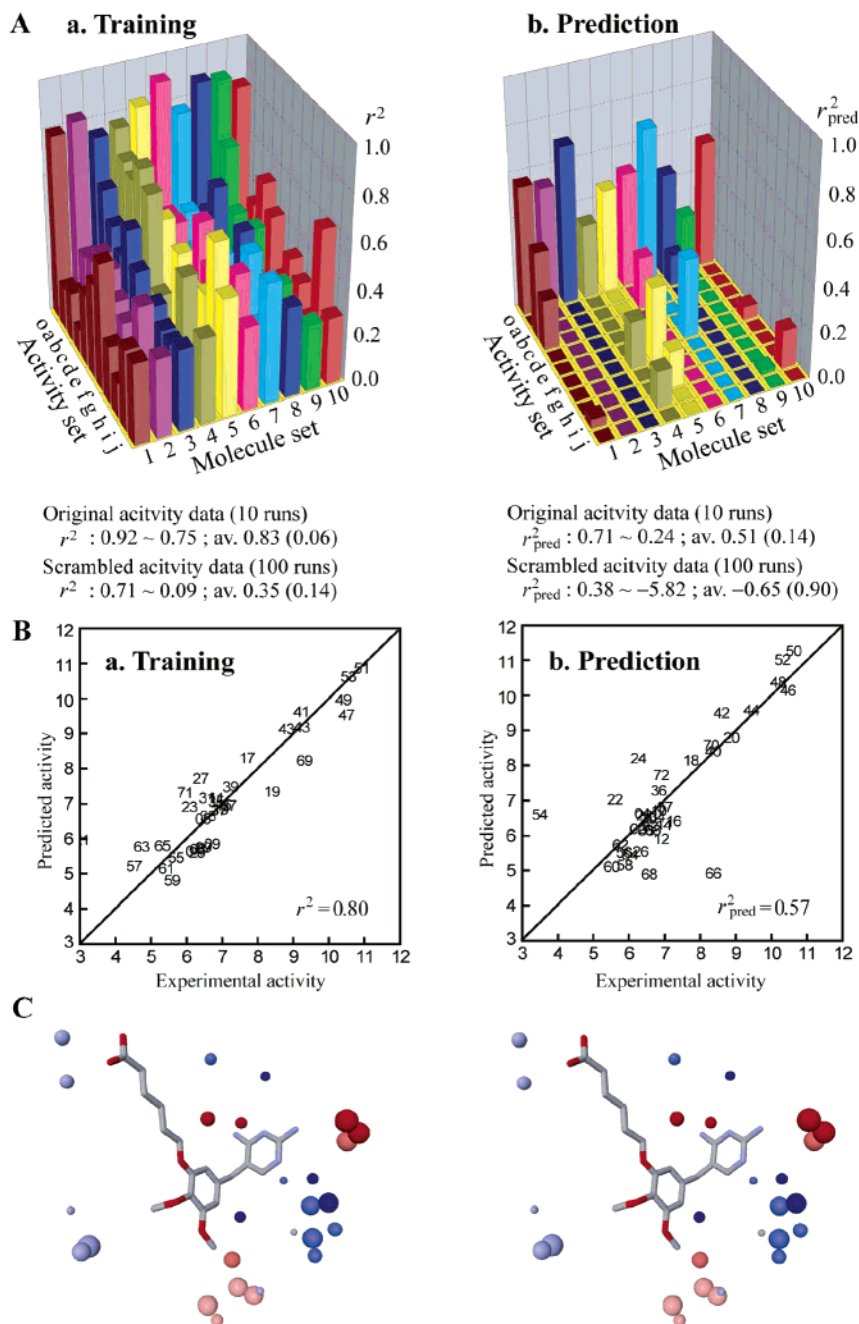
**Table 5.** Structures and Binding Affinities of the DHFR Set

No.	R	$A_{obs}^a$	No.	R	$A_{obs}^a$
		1-53			54-65
1	H	6.18	25	3-CH <sub>2</sub> OH	6.28
2	3-F	6.23	26	3,5-(CH <sub>2</sub> OH) <sub>2</sub>	6.31
3	4-NH <sub>2</sub>	6.30	27	3-O(CH <sub>2</sub> ) <sub>6</sub> CH <sub>3</sub>	6.39
4	4-F	6.35	28	4-OCH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	6.40
5	4-Cl	6.45	29	3-OH	6.47
6	3,4-(OH) <sub>2</sub>	6.46	30	3-OCH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	6.53
7	3-CH <sub>3</sub>	6.70	31	3-CH <sub>2</sub> O(CH <sub>2</sub> ) <sub>3</sub> OCH <sub>3</sub>	6.55
8	3-Cl	6.65	32	3-OCH <sub>2</sub> CONH <sub>2</sub>	6.57
9	3-CH <sub>3</sub>	6.70	33	3-CH <sub>2</sub> OCH <sub>3</sub>	6.59
10	4-Br	6.82	34	4-N(CH <sub>3</sub> ) <sub>2</sub>	6.78
11	4-OH <sub>3</sub>	6.82	35	3-O(CH <sub>2</sub> ) <sub>6</sub> CH <sub>3</sub>	6.82
12	4-NHCOCH <sub>3</sub>	6.89	36	3-O(CH <sub>2</sub> ) <sub>6</sub> CH <sub>3</sub>	6.82
13	3-OCH <sub>3</sub>	6.93	37	4-O(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>	6.89
14	3-Br	6.96	38	4-OCH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	6.99
15	3-CH <sub>3</sub>	7.02	39	3,4-(OCH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub> ) <sub>2</sub>	7.22
16	3-I	7.23	40	3,5-(OCH <sub>2</sub> ) <sub>2</sub> -4-O(CH <sub>2</sub> ) <sub>2</sub> OCH <sub>3</sub>	8.35
17	3-CF <sub>3</sub> -4-OCH <sub>3</sub>	7.69	41	3,5-(OCH <sub>2</sub> ) <sub>2</sub> -4-Br	9.22
18	3,4-(OCH <sub>3</sub> ) <sub>2</sub>	7.72	42	3,4-(OCH <sub>3</sub> ) <sub>2</sub> -5-OCH <sub>2</sub> CO <sub>2</sub> H	8.59
19	3,5-(OCH <sub>3</sub> ) <sub>2</sub>	8.38	43	3-OCH <sub>3</sub> -4-Br-5-OCH <sub>2</sub> CO <sub>2</sub> H	8.80
20	3,4,5-(OCH <sub>3</sub> ) <sub>3</sub>	8.87	44	3,4-(OCH <sub>3</sub> ) <sub>2</sub> -5-O(CH <sub>2</sub> ) <sub>3</sub> CO <sub>2</sub> H	9.43
21	3,5-(OH) <sub>2</sub>	3.04	45	3-OCH <sub>3</sub> -4-Br-5-O(CH <sub>2</sub> ) <sub>2</sub> CO <sub>2</sub> H	9.23
22	4-O(CH <sub>2</sub> ) <sub>6</sub> CH <sub>3</sub>	5.60	46	3,4-(OCH <sub>3</sub> ) <sub>2</sub> -5-O(CH <sub>2</sub> ) <sub>3</sub> CO <sub>2</sub> H	10.46
23	4-O(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>	6.07	47	3,4-(OCH <sub>3</sub> ) <sub>2</sub> -5-O(CH <sub>2</sub> ) <sub>3</sub> CO <sub>2</sub> H	10.49
24	3-O(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>	6.25	48	3,4-(OCH <sub>3</sub> ) <sub>2</sub> -5-O(CH <sub>2</sub> ) <sub>4</sub> CO <sub>2</sub> H	10.18
49	3-OCH <sub>3</sub> -4-Br-5-O(CH <sub>2</sub> ) <sub>3</sub> CO <sub>2</sub> H	10.40	58	3-Cl	5.87
50	3,4-(OCH <sub>3</sub> ) <sub>2</sub> -5-O(CH <sub>2</sub> ) <sub>3</sub> CO <sub>2</sub> H	10.62	59	3-I	5.58
51	3-OCH <sub>3</sub> -4-Br-5-O(CH <sub>2</sub> ) <sub>3</sub> CO <sub>2</sub> H	10.92	60	3-CN	5.51
52	3,4-(OCH <sub>3</sub> ) <sub>2</sub> -5-O(CH <sub>2</sub> ) <sub>3</sub> CO <sub>2</sub> H	10.30	61	3-CH <sub>3</sub>	5.42
53	3,4-(OCH <sub>3</sub> ) <sub>2</sub> -4-Br-5-O(CH <sub>2</sub> ) <sub>6</sub> CO <sub>2</sub> H	10.54	62	3-(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>	5.75
54	3-CONH <sub>2</sub>	3.48	63	3-C(CH <sub>3</sub> ) <sub>3</sub>	4.72
55	3-CF <sub>3</sub>	5.69	64	3-O(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>	6.02
56	3-F	5.85	65	3-OCH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	5.31
57	H	4.51			
		66			67
66		8.36	68	(pyrimethamine)	6.55
67		7.17			
		69,70			71,72
69	3,5-(OCH <sub>3</sub> ) <sub>2</sub>	9.31	71	3,5-(OCH <sub>3</sub> ) <sub>2</sub>	5.95
70	4-OCH <sub>3</sub>	8.30	72	4-OCH <sub>3</sub>	6.89

<sup>a</sup>  $A_{obs}$  is the experimental activity in p*K*<sub>i</sub>.

around the ligands that contains no surviving probe. This region is exactly coincident with the solvent-exposed region of the protein. The result for the DHFR case is indicative of

a possibility that the WeP method may be able to produce the pseudoreceptors that mimic the real receptor binding surface to a certain extent.



**Figure 4.** (A) Plots of the correlation coefficients with other statistical values for (a) training ( $r^2$ ) and (b) prediction ( $r^2_{pred}$ ) obtained from the permutation and scramble tests. See the legends in Figure 2(C) for the detailed descriptions of the plots. (B) Experimental versus predicted binding affinities for (a) training and (b) prediction of the representative DHFR set. (C) Stereoview of the pseudoreceptor for the representative DHFR set. The most active compound **24** was overlaid as a frame of reference. See the legends in Figure 2(B) for the detailed descriptions of the spheres.

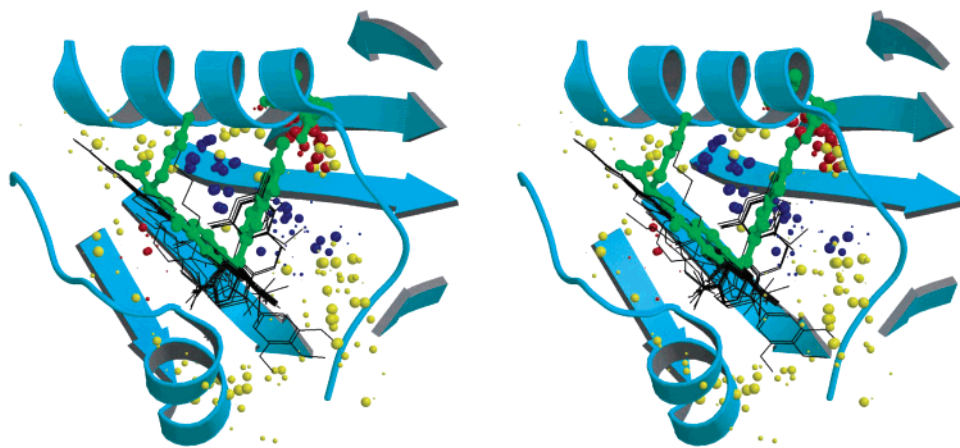
## DISCUSSION

During the development of the present method, we had to settle several issues that were related either to the basic formulations in the methodology or to the working procedures. The procedures described and the parameters used herein reflect the results of numerous experiments to optimize these interlinked factors.

The most important issues were what kinds of atom type would be used for the probes and interaction energy terms and how to assign the probe charges. We used only the van der Waals and electrostatic interaction terms for the probe-ligand interaction energies, since the incorporation of the hydrogen bonding energies was not straightforward in the

current scheme. Without any information on the distribution of partial charges in the receptor, we had to resort to the electrostatic potentials of the ligands to somehow assign the charges to the surface probes. If the optimal probe charges were to be determined a priori, the number of unknowns was doubled, which is quite undesirable in a problem highly underdetermined per se. In fact, we attempted to refine the probe charges after the initial convergence, but there was essentially no improvement in the final correlations. We have found that, given the probe charges, the overall optimization behavior and the final distributions of the probe weights were relatively insensitive to the probe atom type. We have confirmed that the effects of the separations between the





**Figure 5.** Stereoview of the DHFR-folate complex and the WeP pseudoreceptors. All of the surviving surface probes from the 10 permutation runs together with the aligned ligands were superposed onto the crystal structure of the human DHFR-folate complex (1dhf). Each sphere representing the surviving probes has a radius proportional to the weight and is colored in yellow when the probe is neutral (partial charge:  $-0.2 \sim +0.2$ ), in red when electronegative ( $< -0.2$ ), or in blue when electropositive ( $> +0.2$ ). DHFR was drawn schematically in ribbon only for the region that interacts with the ligands, and its side chains were omitted for clarity. Note that the conformations of the aligned ligands (denoted in black lines) were energy minimized in the absence of the protein and thus are different from that of the folate (denoted with green-colored ball-and-stick) in complex with the protein.

probes and their absolute positions were also negligible. For each of the three test cases, the surface probes were generated with grid spacings of 1.0, 1.5, and 2.0 Å using the marching cube algorithm, and the WeP analyses were performed. As the grid size became smaller, the number of surviving probes increased significantly, not to mention the computing time, but the correlation improved only marginally. Therefore a grid spacing of 2.0 Å was taken as the default value to generate the surface probes. When the initial grids were shifted by half the grid size along the three axes in the Cartesian coordinate system, the results remained essentially unchanged indicating that the WeP analyses do not depend on the absolute positions of probe sites.

Our experiments indicated that, under the present WeP scheme, decent results could be obtained even with a simplistic approach in the energy evaluation. In this regard, our WeP method sharply contrasts with other receptor surface methods in which various types of probes or sophisticated energy evaluation schemes are utilized trying to evaluate the ligand binding energy as accurately as possible.<sup>28,30,37</sup> In the WeP method, all probes have an equal atom type of the methyl group, and each of them has a specific partial charge complementary to the electrostatic potential of the ligands. The interaction energies do not include hydrogen bonding energy and hydrophobic energy terms. It is noteworthy that our results for both hydrophilic DHFR and hydrophobic AHR compounds are still quite acceptable. The effects of many factors that might be interlinked influencing the results seem capable of being amalgamated into a single entity, namely, the probe weights. This feature may be an important asset of the WeP approach. In most 3D-QSAR methods that utilize the interaction energies, the results critically depend on the selection of the conformers of the ligand molecules and their mutual alignment.<sup>4</sup> The same holds true for the WeP method. We use the united atom force fields as a minimal measure to eliminate the conformational effects due to the H atoms whose positions are often ambiguous.

We have examined the effects of reproduction algorithms and some parameters in GA on the convergence behavior. It is well-known that the use of “elitists” in GA can lead to

a loss of genetic diversity and thus to early convergence that results in a lower fitness score. In this study, however, we still preferred using the elitism for reproduction. We found that the use of elitism, if combined with nonredundancy tests, was an efficient way to maintain genetic diversity through generations, assuredly preventing the models from becoming inbred. Compared to democratic nonelitist strategies, the same level of fitness could be achieved with smaller numbers of populations and generations.

Usually the best population size is both application-dependent and related to the length of the chromosome, and a reasonable compromise between these two parameters should be adopted based on empirical experimentation.<sup>45</sup> In this study, population sizes of 1000, 2000, and 2000 were used for the AHR, CBG, and DHFR sets, respectively, which corresponded to 9.3, 16.1, and 8.8 times the number of initial probes for their representative sets. Experimentation indicated that the population size was no longer a significant factor in the optimization if it exceeded about five times the number of initial probes. In case of the CBG set, for instance, tests with population sizes of  $2 \times 10^5$  and  $2 \times 10^3$  gave very similar correlations, although not exactly the same due to the stochastic nature of GA. Differences were limited to the probes with very small weight values. In contrast, the correlations were degraded while the models evidently became inbred, if the population sizes smaller than a threshold value were used. From the standpoint of stability of the optimization and computational efficiency, therefore, the optimal population size was estimated to be about 10–20 times the number of the initial probes. To ensure near complete convergence, we used rather stringent criteria. In all three test samples, the systems converged within 1000 generations.

We have performed extensive permutation and scramble tests for all three samples. Comparison of these results indicated that the WeP models did not suffer from overfitting and were not prone to chance correlation as long as the ratio between the number of probes and the number of populations exceeded the threshold value of five. In all permutation tests, the trainings were performed in relatively stable ways. The

number and location of the surviving probes heavily weighted did not vary significantly regardless of the selections of the training molecules, and the variations among different sets were limited to the probes with small weight values in all three cases. It is interesting to note that, in the scramble tests, the average  $r^2$  values for trainings were similar with values of 0.33 ( $\pm 0.16$ ), 0.37 ( $\pm 0.13$ ), and 0.35 ( $\pm 0.14$ ) for the AHR, CBG, and DHFR sets, respectively, while, in the permutation tests, they differed considerably from each other with values of 0.94 ( $\pm 0.03$ ), 0.87 ( $\pm 0.04$ ), and 0.83 ( $\pm 0.05$ ), respectively. These suggest that the trainings with the WeP method would probably result in an average  $r^2$  value around 0.35 for any sample with the activity data randomly selected within certain ranges.

The WeP method has both similarities and dissimilarities compared to other receptor surface methods such as GERM,<sup>30</sup> FLARM,<sup>33,34</sup> and Quasar.<sup>35–37</sup> These methods are similar to each other in that the surface probes are used as a receptor surrogate, the total probe interaction energies instead of the individual field values are correlated with the activities, and the optimizations are performed using GA. However, our method is unique in terms of the variables which are to be optimized. The WeP method tries to find the optimal distribution of the weight values for the surface probes, while other methods try to optimize the atom types or properties. In fact, the GERM and Quasar methods contain a null atom, which does not contribute to the binding energy and thus corresponds to a probe with zero weight in the WeP, as one of the available atom types. In several results using these methods, however, only a fraction of the probes were assigned the null atom upon optimization, so that the pseudoreceptor covers a substantial part of the surface surrounding the aligned ligands.<sup>30,35–37</sup> In contrast, the WeP pseudoreceptors cover a much smaller portion of the receptor surface. In the representative sets of the AHR, CBG, and DHFR cases, the pseudoreceptors comprise only 9% (10 surviving probes from 107 initial probes), 22% (27/124), and 14% (32/228) of the initial probes, respectively. Especially, the results from the DHFR set are encouraging, since a wide region that is totally void of surviving probes coincides with the solvent region of the DHFR protein. Identification of the solvent exposed region through the 3D-QSAR analyses of ligand data alone would be very beneficial to drug design. It remains to be confirmed whether the WeP method always has this capability.

To enhance the predictability of the WeP method, several steps in the current scheme may have to be improved. They may include a more elaborate energy evaluation scheme, the use of a nonlinear relationship between the probe interaction energies and activities, the use of more sophisticated diversity operators in GA, and/or the refinement of probe positions and probe charges in conjunction with the GA optimization. Most importantly, the method should be capable of taking account of the conformational flexibility and molecular alignment. Further studies are deemed worthwhile even though the WeP method seems a working 3D-QSAR technique in its present form.

In summary, the present study shows that a rather simple concept of the weighted probe interaction energy could be formulated into a 3D-QSAR method called the WeP. The WeP pseudoreceptors can be trained in a stable manner, possess relatively high predictive ability, and may simulate

the solvent exposed regions of real proteins. The WeP method can directly be used to guide the design of more potent compounds.

## ACKNOWLEDGMENT

This research was supported by a grant (CBM-02-B-3) from the Center for Biological Modulators of the 21st Century Frontier R&D Program, the Ministry of Science and Technology, Korea.

## REFERENCES AND NOTES

- (1) Kubinyi, H. *3D QSAR in Drug Design: Theory, Methods and Applications*; ESCOM: Leiden, 1993.
- (2) Platt, D. E.; Parida, L.; Gao, Y.; Floratos, A.; Rigoutsos, I. QSAR in grossly underdetermined systems: Opportunities and issues. *IBM J. Res. Dev.* **2001**, *45*, 533–544.
- (3) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- (4) Kubinyi, H. Comparative molecular field analysis (CoMFA). In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R. Eds.; John Wiley & Sons: Chichester, 1998; pp 448–460.
- (5) Norinder, U. Recent progress in CoMFA methodology and related techniques. In *3D QSAR in Drug Design. Volume 2. Ligand-Protein Interactions and Molecular Similarity*; Kubinyi, H., Folkers, G., Martin, Y. C. Eds.; Kluwer/ESCOM: Dordrecht, 1998; pp 25–39.
- (6) Bush, B. L.; Nachbar, R. B. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587–619.
- (7) Cruciani, G.; Watson, K. A. Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. *J. Med. Chem.* **1994**, *37*, 2589–2601.
- (8) Cho, S. J.; Tropsha, A. Cross-validated  $R^2$ -guided region selection for comparative molecular field analysis: A simple method to achieve consistent results. *J. Med. Chem.* **1995**, *38*, 1060–1066.
- (9) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B. Q.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR models using 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524.
- (10) Pastor, M.; Cruciani, G.; Clementi, S. Smart region definition (SRD): A new way to improve the predictive ability and interpretability of 3D-QSAR models. *J. Med. Chem.* **1997**, *40*, 1455–1464.
- (11) Wold, S.; Johansson, E.; Cocci, M. PLS-partial least squares projection to latent structures. In *3D QSAR in Drug Design*; Kubinyi, H. Ed.; ESCOM: Leiden, 1993; pp 523–563.
- (12) Kim, K. H.; Greco, G.; Novellino, E.; Silipo, C.; Vittoria, A. Use of the hydrogen bond potential function in a comparative molecular field analysis (CoMFA) on a set of benzodiazepines. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 263–280.
- (13) Leo, A. J.; Hansch, C. Role of hydrophobic effects in mechanistic QSAR. *J. Comput.-Aid. Mol. Des.* **1999**, *17*, 1–25.
- (14) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.
- (15) Gebauer, S.; Knutter, I.; Hartrodt, B.; Brandsch, M.; Neubert, K.; Thondorf, I. Three-dimensional quantitative structure–activity relationship analyses of peptide substrates of the mammalian H<sup>+</sup>/peptide cotransporter PEPT1. *J. Med. Chem.* **2003**, *46*, 5725–5734.
- (16) Bostrom, J.; Bohm, M.; Gundertofte, K.; Klebe, G. A 3D QSAR study on a set of dopamine D<sub>4</sub> receptor antagonists. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1020–1027.
- (17) Kuo, C.-L.; Assefa, H.; Kamath, S.; Brzozowski, Z.; Slawinski, J.; Saczewski, F.; Buolamwini, J. K.; Neamati, N. Application of CoMFA and CoMSIA 3D-QSAR and docking studies in optimization of mercaptobenzenesulfonamides as HIV-1 integrase inhibitors. *J. Med. Chem.* **2004**, *47*, 385–399.
- (18) Kubinyi, H.; Hamprecht, F. A.; Mietzner, T. Three-dimensional quantitative similarity-activity relationships (3D QSAR) from SEAL similarity matrices. *J. Med. Chem.* **1998**, *41*, 2553–2564.
- (19) So, S.-S.; Karplus, M. Three-dimensional quantitative structure–activity relationships from molecular similarity matrices and genetic neural networks. 1. Method and validations. *J. Med. Chem.* **1997**, *40*, 4347–4359.

- (20) So, S.-S.; Karplus, M. Three-dimensional quantitative structure activity relationships from molecular similarity matrices and genetic neural networks. 2. Applications. *J. Med. Chem.* **1997**, *40*, 4360–4371.
- (21) Doweiko, A. M. The hypothetical active site lattice. An approach to modelling active site from data on inhibitor molecules. *J. Med. Chem.* **1988**, *31*, 1396–1406.
- (22) Robinson, D. D.; Winn, P. J.; Lyne, P. D.; Richards, W. G. self-organizing molecular field analysis: A tool for structure–activity studies. *J. Med. Chem.* **1999**, *42*, 573–583.
- (23) Jain, A. N.; Koile, K.; Chapman, D. Compass: Predicting biological activities from molecular surface properties. Performance comparisons on a steroid benchmark. *J. Med. Chem.* **1994**, *37*, 2315–2327.
- (24) Wagener, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of molecular surface properties for modeling corticosteroid binding globulin and cytosolic Ah receptor activity by neural networks. *J. Am. Chem. Soc.* **1995**, *117*, 7769–7775.
- (25) Bravi, G.; Gancia, E.; Mascagni, P.; Pegna, M.; Todeschini, R.; Zaliani, A. MS–WHIM, new 3D theoretical descriptors derived from molecular surface properties: A comparative 3D QSAR study in a series of steroids. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 79–92.
- (26) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRIND-INdependent Descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233–3243.
- (27) Polanski, J.; Giełeciak, R.; Bak, A. The comparative molecular surface analysis (COMSA) – A nongrid 3D QSAR method by a coupled neural network and PLS system: Predicting pKa values of benzoic and alkanolic acids. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 184–191.
- (28) Hahn, M. Receptor surface models. 1. Definition and construction. *J. Med. Chem.* **1995**, *38*, 2080–2090.
- (29) Hahn, M.; Rogers, D. Receptor surface models. 2. Application to quantitative structure–activity relationships studies. *J. Med. Chem.* **1995**, *38*, 2091–2102.
- (30) Walters, D. E.; Hinds, R. M. Genetically evolved receptor models (GERM): A computational approach to construction of receptor models. *J. Med. Chem.* **1994**, *37*, 2527–2536.
- (31) Chen, H.; Zhou, J.; Xie, G. PARM: A genetic evolved algorithm to predict bioactivity. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 243–250.
- (32) Pei, J.; Zhou, J.; Xie, G.; Chen, H.; He, X. PARM: A practical utility for drug design. *J. Mol. Graph. Model.* **2001**, *19*, 448–454.
- (33) Pei, J.; Zhou, J.; Xie, G.; Chen, H.; He, X. Flexible atom receptor model (FLARM). *Acta Chim. Sinica* **2002**, *60*, 973–979.
- (34) Peng, T.; Pei, J.; Zhou, J. 3D-QSAR and receptor modeling of tyrosine kinase inhibitors with flexible atom receptor model (FLARM). *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 298–303.
- (35) Vedani, A.; Dobler, M.; Zbinden, P. Quasi-atomistic receptor surface models: A bridge between 3-D QSAR and receptor modeling. *J. Am. Chem. Soc.* **1998**, *120*, 4471–4477.
- (36) Vedani, A.; Briem, H.; Dobler, M.; Dollinger, H.; McMasters, D. R. Multiple-conformation and protonation-state representation in 4D-QSAR: The neurokinin-1 receptor system. *J. Med. Chem.* **2000**, *43*, 4416–4427.
- (37) Vedani, A.; Dobler, M. 5D-QSAR: The key for simulating induced fit? *J. Med. Chem.* **2002**, *45*, 2139–2149.
- (38) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.* **1995**, *38*, 2681–2691.
- (39) Wang, T.; Wade, R. C. Comparative binding energy (COMBINE) analysis of influenza neuraminidase-inhibitor complexes. *J. Med. Chem.* **2001**, *44*, 961–971.
- (40) Lorensen, W. E.; Cline, H. E. Marching cubes: A high resolution 3D surface construction algorithm. *Computer Graphics* **1987**, *21*, 163–169.
- (41) Heiden, W.; Goetze, T.; Brickmann, J. Fast generation of molecular surfaces from 3D data fields with an enhanced “marching cube” algorithm. *J. Comput. Chem.* **1993**, *14*, 246–250.
- (42) Gasteiger, H.; Marsili, M. Iterative partial equalization of orbital electronegativity – A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3288.
- (43) Shin, W.; Oh, D.-G.; Chae, C. H.; Yoon, T. S. Conformational analysis of thiamin-related compounds: A stereochemical model for thiamin catalysis. *J. Am. Chem. Soc.* **1993**, *115*, 12238–12250.
- (44) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Salvatore, P., Jr.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
- (45) Judson, R. Genetic algorithms and their use in chemistry. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1997; pp 1–73.
- (46) Kubinyi, H.; Abraham, U. Practical problems in PLS analysis. In *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H. Ed.; ESCOM: Leiden, 1993; pp 717–728.
- (47) Safe, S. Polychlorinated biphenyls (PCBs), dibenzo-*p*-dioxins (PCDDs), dibenzofurans (PCDFs), and related compounds: Environmental and mechanistic considerations which support the development of toxic equivalency factors (TEFs). *Crit. Rev. Toxicol.* **1990**, *21*, 51–88.
- (48) Coats, E. A. The CoMFA steroids as a benchmark dataset for development of 3D QSAR methods. In *3D QSAR in Drug Design. Volume 3. Recent Advances*; Kubinyi, H., Folkers, G., Martin, Y. C., Eds.; Kluwer/ESCOM: Dordrecht, 1998; pp 199–213.
- (49) <http://www2.ccc.uni-erlangen.de/services/steroids/index.html>.
- (50) Davies, J. F., II.; Delcamp, T. J.; Prendergast, N. J.; Ashford, V. A.; Freisheim, J. H.; Kraut, J. Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafolate. *Biochemistry* **1990**, *29*, 9467–9479.

CI0498721