# Prediction-Weighted Partial Least-Squares Regression Method (PWPLS) 2: Application to CoMFA

Yukio Tominaga* and Iwao Fujiwara

Department of Chemistry I, Discovery Research Laboratories I, Dainippon Pharmaceutical Co., Ltd. Enoki 33−94, Suita/Osaka 564, Japan

Received April 18, 1997[⊗]

Comparative molecular field analysis (CoMFA) has been used in drug design and three-dimensional quantitative structure−activity relationships (3D-QSAR). In CoMFA analysis, Partial Least-Squares (PLS) is used to correlate a large number of variables with biological activity. However, PLS may not clearly indicate which variables affect the biological activity of compounds. We have developed PWPLS (Prediction-Weighted Partial Least-Squares) that can select good predictor variables and weight each predictor variable to improve the predictiveness of its model. In addition to PWPLS, we developed another method, $Q^2$ oriented variable selection (QOVS), to select variables that also affect predictiveness of its model. In this paper, we applied PWPLS and QOVS to the CoMFA study reported by Dunne et al. for binding 21 steroids to corticosteroid-binding globulin (CBG).

## INTRODUCTION

Since the study of Cramer[1] was published, CoMFA has been used widely in drug design and 3D-QSAR.[2−7] CoMFA provides a large number of variables describing nonbonded steric and/or electrostatic interaction energies between probes and each molecule and uses PLS[8−10] to correlate the variables to the observed biological activities.

Using PLS, however, it may be difficult to obtain which variables affect the biological activity of compounds. For instance, a model derived from PLS may not clearly indicate the variables affecting the predictiveness of the model because some variables may provide only subtle effect (noise) on the final model. It has been demonstrated that, while a large number of variables in a model can produce determinal effects on the predictiveness of a model, those variables could in fact contain such noise.[11,12]

What is the noise of CoMFA analysis? In CoMFA analysis, each molecule is located within fixed lattice points. So there is an assumption that each molecule interacts with target receptor for all the directions, that is whole the molecule interacts with the receptor. Recent X-ray crystallography of protein−ligand complex, however, has revealed that the whole parts of the molecule do not always interact with the receptor; in other words, some parts of the molecule do not interact with receptor.[13] These parts are considered to be the noise of CoMFA analysis. Therefore, a method should be established to successfully select the variables with the most significant effect on the biological activity.

We have developed a simple and efficient variable selection method called PWPLS (Prediction-Weighted Partial Least-Squares).[14] PWPLS consists of two steps: the first step is to select variables that affect the predictiveness of a model; in other words, the weight vector of eliminated variables is set to zero; and the second step is PLS analysis by using a prediction-weight vector. In addition to PWPLS, we developed another method, $Q^2$ oriented variable selection (QOVS), to select variables that also affect predictiveness of the model.

In this paper, we applied PWPLS and QOVS to the CoMFA study by Dunne et al.[15] that uses a data set of 21 steroids with corticosteroid-binding globulin (CBG) presented, and compared the results of PWPLS and QOVS with those of PLS for their predictiveness.

## METHOD

**1. Data Set.** The data set is 21 steroids to CBG that has been reported by Dunn et al. CoMFA fields were taken directly from the SYBYL demonstration files that were presented by Cramer et al.[1] Steric and electrostatic fields on a 2 Å grid provided a total of 800 explanatory variables per compound. All calculations used option MINIMUN_SIGMA = 0.0, SCALING = NONE.

**2. PWPLS.** PWPLS consists of two steps: the first step is to select variables that affect predictiveness of the model; in other words, the weight vector of eliminated variables is set to zero; and the second step is PLS analysis by using the predictive weight vector.

**2.1. Selection of Variables.** In PLS1, the correlation vector between **X** (variables) and **y** (biological activity) is used as the weight vector **w** (eq 1) to extract the information from **X**.

$$\mathbf{w} = \mathbf{X}'\mathbf{y}/|\mathbf{X}'\mathbf{y}| \tag{1}$$

This method is widely used to improve the fit between **X** and **y**. However, it is important to consider not only the fit but also the predictiveness because even if their correlation coefficients are equal, the predictiveness, e.g., cross-validated multiple correlation coefficient of predictions ($Q^2$) from leave-one-out method, is not necessarily equal. We therefore determined $Q^2$ of simple-least squares between each variable and biological activity to judge if a variable affects the predictiveness of the model. For example, if the $Q^2$ between a variable and biological activity is less than zero, the variable is excluded from the analysis. This process is then repetitively applied to all variables. This method eliminates the random noise variables showing low $R^2$ and $Q^2$ as well as the variables with high $R^2$ but low $Q^2$. In this study the

[⊗] Abstract published in *Advance ACS Abstracts,* October 15, 1997.

S0095-2338(97)00025-5 CCC: $14.00 

**Table 1.** CoMFA Statistics for 21 Steroids to Corticosteroid-Binding Globulin.

| | optimal no. of components | SDEP[a] | $Q^2$ | SDEC[b] | $R^2$ |
|---|---|---|---|---|---|
| steric field | 2 | 0.606 | 0.719 | 0.374 | 0.893 |
| electrostatic field | 4 | 0.555 | 0.764 | 0.220 | 0.962 |

[a] Standard deviation of error of predictions[18] SDEP = $[\Sigma(y - y_{pred})^2/N]^{1/2}$. [b] Standard deviation of error of calculations[18] SDEC = $[\Sigma(y - y_{calc})^2/N]^{1/2}$.

**Table 2.** PWPLS Analysis of 21 Steroids to Corticosteroid-Binding Globulin Using Steric field[a]

| criteria | no. of grid points | optimal no. of components | SDEP | $Q^2$ | SDEC | $R^2$ |
|---|---|---|---|---|---|---|
| criterion I ($Q^2 > 0.0$) | 445 | 4 | 0.592 | 0.732 | 0.441 | 0.851 |
| criterion II ($Q^2 > 0.1$) | 377 | 2 | 0.570 | 0.751 | 0.447 | 0.846 |
| criterion III ($Q^2 > 0.2$) | 314 | 3 | 0.555 | 0.764 | 0.439 | 0.852 |
| criterion IV ($Q^2 > 0.3$) | 237 | 3 | 0.600 | 0.724 | 0.444 | 0.849 |
| criterion V ($Q^2 > 0.4$) | 149 | 1 | 0.522 | 0.792 | 0.447 | 0.847 |
| criterion VI ($Q^2 > 0.5$) | 9 | 1 | 0.589 | 0.734 | 0.532 | 0.783 |

[a] The table entries report the values of SDEP, $Q^2$, SDEC, and $R^2$ for each criteria.

**Table 3.** PWPLS Analysis of 21 Steroids to Corticosteroid-Binding Globulin Using Electrostatic Field[a]

| criteria | no. of grid points | optimal no. of components | SDEP | $Q^2$ | SDEC | $R^2$ |
|---|---|---|---|---|---|---|
| criterion I ($Q^2 > 0.0$) | 376 | 6 | 0.603 | 0.722 | 0.336 | 0.913 |
| criterion II ($Q^2 > 0.1$) | 284 | 6 | 0.595 | 0.729 | 0.313 | 0.924 |
| criterion III ($Q^2 > 0.2$) | 188 | 3 | 0.545 | 0.773 | 0.422 | 0.863 |
| criterion IV ($Q^2 > 0.3$) | 98 | 2 | 0.571 | 0.750 | 0.850 | 0.442 |
| criterion V ($Q^2 > 0.4$) | 18 | 4 | 0.680 | 0.647 | 0.816 | 0.490 |
| criterion VI ($Q^2 > 0.5$) | 2 | 1 | 0.713 | 0.611 | 0.622 | 0.704 |

[a] The table entries report the values of SDEP, $Q^2$, SDEC, and $R^2$ for each criteria.

following six criteria were determined: $Q^2$ as less than 0.0 (criterion I), 0.1 (criterion II), 0.2 (criterion III), 0.3 (criterion IV), 0.4 (criterion V), and 0.5 (criterion VI). This is the first stage of PWPLS.

**2.2. Prediction-Weights.** In PLS1, the fit is only the parameter considered to extract the information of variables even though it does not always correlate with the predictiveness. So we introduced a factor of predictiveness when the information is extracted from variables. We then redefined the weight vector to the vector consisting of the product of the correlation matrix **X′y** and predictiveness factor **PF**.

$$\mathbf{w} = \mathbf{X'yPF}/d \quad d = |\mathbf{X'yPF}| \tag{2}$$

$Q^2$ is not always appropriate for **PF** because $Q^2$ may contain a negative value when the predictiveness is low. We thus defined the predictive correlation coefficient **PR** as follows: **PR** is the correlation coefficient between observed **y** and predicted **y** that was estimated from simple least-square regression by cross-validation with the leave-one-out method.

**Table 4.** CoMFA Analysis of 21 Steroids to Corticosteroid-Binding Globulin Using Steric Field[a]

| criteria | no. of grid points | optimal no. of components | SDEP | $Q^2$ | SDEC | $R^2$ |
|---|---|---|---|---|---|---|
| criterion A ($dQ^2 > 0.00$) | 51 | 9 | 0.428 | 0.860 | 0.242 | 0.955 |
| criterion B ($dQ^2 > 0.01$) | 8 | 8 | 0.420 | 0.865 | 0.302 | 0.930 |
| criterion C ($dQ^2 > 0.02$) | 5 | 4 | 0.474 | 0.828 | 0.391 | 0.882 |
| criterion D ($dQ^2 > 0.03$) | 3 | 3 | 0.483 | 0.821 | 0.434 | 0.855 |
| criterion E ($dQ^2 > 0.04$) | 4 | 4 | 0.404 | 0.875 | 0.457 | 0.840 |
| criterion F ($dQ^2 > 0.05$) | 2 | 2 | 0.527 | 0.787 | 0.569 | 0.752 |
| criterion G ($dQ^2 > 0.10$) | 1 | 1 | 0.589 | 0.734 | 0.713 | 0.612 |

[a] The table entries report the values of SDEP, $Q^2$, SDEC, and $R^2$ for each criteria.

**Table 5.** CoMFA Analysis of 21 Steroids to Corticosteroid-Binding Globulin Using Electrostatic Field[a]

| criteria | no. of grid points | optimal no. of components | SDEP | $Q^2$ | SDEC | $R^2$ |
|---|---|---|---|---|---|---|
| criterion A ($dQ^2 > 0.00$) | 49 | 13 | 0.279 | 0.940 | 0.127 | 0.987 |
| criterion B ($dQ^2 > 0.01$) | 16 | 12 | 0.258 | 0.949 | 0.131 | 0.986 |
| criterion C ($dQ^2 > 0.02$) | 3 | 3 | 0.548 | 0.770 | 0.655 | 0.672 |
| criterion D ($dQ^2 > 0.03$) | 3 | 3 | 0.542 | 0.775 | 0.655 | 0.672 |
| criterion E ($dQ^2 > 0.04$) | 3 | 3 | 0.535 | 0.781 | 0.655 | 0.672 |
| criterion F ($dQ^2 > 0.05$) | 2 | 2 | 0.583 | 0.740 | 0.659 | 0.668 |
| criterion G ($dQ^2 > 0.10$) | 2 | 2 | 0.583 | 0.740 | 0.659 | 0.668 |

[a] The table entries report the values of SDEP, $Q^2$, SDEC, and $R^2$ for each criteria.

**Table 6.** CoMFA Analysis of 21 Steroids to Corticosteroid-Binding Globulin Using Steric Field or Electrostatic Field[a]

| method | field type | no. of grid points | optimal no. of components | $Q^2$ | $R^2$ |
|---|---|---|---|---|---|
| PLS | steric field | 800 | 2 | 0.719 | 0.893 |
| PLS | electrostatic field | 800 | 4 | 0.764 | 0.962 |
| PWPLS | steric field | 149 | 1 | 0.792 | 0.847 |
| PWPLS | electrostatic field | 188 | 3 | 0.773 | 0.863 |
| QOVS | steric field | 4 | 4 | 0.875 | 0.840 |
| QOVS | electrostatic field | 16 | 3 | 0.807 | 0.892 |

[a] The table entries report the values of $Q^2$ and $R^2$ for each method.

$$\mathbf{w} = \mathbf{X'yPR}/d \quad d = |\mathbf{X'yPR}| \tag{3}$$

By using this weight vector, the PWPLS algorism is defined.

**3. $Q^2$ Oriented Variable Selection (QOVS).** We assumed the model equation showing high predictiveness should have used highly predictable variables. To implement this assumption, we first evaluated $Q^2$ for each variable with biological data and then selected the variable showing the highest $Q^2$. Using this variable the initial PLS model was made. This model is equivalent to simple least-square regression. The variable showing the secondly high $Q^2$ was then added to the initial model, and PLS analysis was carried out. If $Q^2$ of this model is higher than that with the initial model, the other variable with the thirdly high $Q^2$ is added

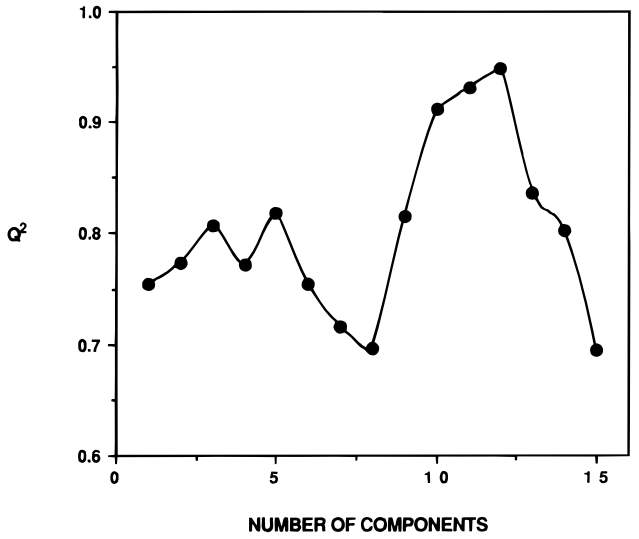


**Figure 1.** The plot of $Q^2$ versus the number of components.

to the model. But if the predictiveness is not improved from the initial model, the variable with the secondary high $Q^2$ is replaced with the other variable with the thirdly high $Q^2$. This process was repetitively applied for all variables. We determined the following seven criteria to judge if the predictiveness of this model has improved: $Q^2$ of the model increased by 0.00 (criterion A), 0.01 (criterion B), 0.02 (criterion C), 0.03 (criterion D), 0.04 (criterion E), 0.05 (criterion F), and 0.10 (criterion G) from the previous model. As a result, the model which shows the best $Q^2$ for this method was obtained.

**4. Data Analysis Method.** All PLS and PWPLS calculations were performed by a self-written FORTRAN-77 program on Indigo 2 running the IRIX Version 6.2. Results of PLS were cross-checked with SYBYL-PLS results.[16,17]

## RESULTS AND DISCUSSIONS

**1. Statistics. 1.1. Normal PLS.** PLS analysis for steric field (800 grid points) and electrostatic field (800 grid points) were carried out separately. The results of steric field and electrostatic field with their full grid points were summarized in Table 1. For steric field, the optimum component is 2, $Q^2$ with leave-one-out method is 0.719, and $R^2$ is 0.893. For electrostatic field, the optimum component is 4, $Q^2$ is 0.764, and $R^2$ is 0.963.

**1.2. PWPLS.** Each $Q^2$ for simple regression was estimated between the biological activity and each grid point of the CoMFA field. The highest $Q^2$ of 0.734 for steric field and 0.596 for electrostatic field were provided by a single grid point. It was surprising to see that only a single grid point of steric field provided superior $Q^2$ (0.734), while $Q^2$ of steric field of 0.719 was obtained with full grid points. PWPLS was then carried out using criterion I ($Q^2 < 0.0$), criterion II ($Q^2 < 0.1$), criterion III ($Q^2 < 0.2$), criterion IV ($Q^2 < 0.3$), criterion V ($Q^2 < 0.4$), and criterion VI ($Q^2 < 0.5$). The number of selected grid points and the statistical results are summarized in Tables 2 and 3. For steric field, the reduction of the variables by criteria 1, 2, 3, 4, 5, and 6 improved the predictiveness as follows: $Q^2$ improved from 0.719 (800 grid points) to 0.732 (criterion I: 445 grid points), to 0.751 (criterion II: 377 grid points), to 0.764 (criterion III: 314 grid points), to 0.724 (criterion IV: 237 grid points), to 0.792 (criterion V: 149 grid points), and to 0.734 (criterion VI: 9 grid points). The best model ($Q^2 = 0.792$, $R^2 = 0.847$) was obtained using criterion V with one component. For electrostatic field, reduction of variables did not always improve the predictiveness. Only criterion III improved $Q^2$ from 0.722 (800 grid points) 0.773 (188 grid points). For this field, the best model ($Q^2 = 0.773$, $R^2 = 0.863$) was obtained using criterion III with three components.

**1.3. QOVS.** The grid point that provided the highest predictiveness for steric and electrostatic fields was used as their initial points. Criteria A to G were used to develop the respective PLS model. The results of steric field are shown in Table 4. The most predictable model for this field was constructed from criterion E with four components; $Q^2$ is 0.875, and $R^2$ is 0.840. $Q^2$ of this model has improved from that obtained with full variables. On the other hand, the fit ($R^2$) of this model did not improve. In this model, the optimum component is equivalent to the number of the reduced variables. This PLS model is thus equivalent to the multiple linear regression model. For electrostatic field, the results are summarized in Table III. The most predictable model was constructed by using criterion B. The optimum component was 12 with obtained $Q^2$ of 0.949 and $R^2$ of 0.986. Although this $Q^2$ is extremely high, the number of components is also relatively high. Figure 1 shows the plots between $Q^2$ and each component that were derived from the PLS analysis using the variables reduced by criterion B. Three optimum $Q^2$s are indicated in the plots, with 12 components ($Q^2 = 0.949$), 5 components ($Q^2 = 0.818$), and 3 components ($Q^2 = 0.807$). To avoid overfitting (or over-inner prediction), we selected the three component model. The model shows $Q^2$ of 0.807 and $R^2$ of 0.892.

**1.4. Comparison of Three Methods.** The statistics of the final model of each method is summarized in Table IV. Among the three methods, QOVS for steric and electrostatic fields shows the best predictiveness (highest $Q^2$). The fit ($R^2$) of QOVS, however, did not improve from that with normal PLS. In these QOVS models, four grid points for steric field and 16 for electrostatic field were selected, and the most predictable PLS model tended to be a full model like multiple linear regression model. The reason is that, in this method, the full information (all the variance) of each independent variable is used to select the variables. In other words, contribution of each variable to the predictiveness of PLS models are estimated from the predictiveness of each independent variable with full variance.

In PWPLS, although the improvement of the predictiveness from that of normal PLS was not good when compared with QVOS, the predictiveness for both steric and electrostatic fields were improved. The fit did not improve, however, from that of normal PLS. Using selected grid points, normal PLS analysis was carried out. The results are as follows: for steric field, the optimum component is one, $Q^2$ is 0.799, and $R^2$ is 0.844. For electrostatic field, optimum component is three, $Q^2$ is 0.741, and $R^2$ is 0.845. The results of PWPLS and PLS with selected variables are almost equivalent. So for the CoMFA data sets used in this study, PWPLS showed significance for the variable selection process but did not indicate such importance for the weighing process.

**2. CoMFA Fields.** The CoMFA steric and electrostatic fields for PLS and PWPLS analyses are shown as contour plots in Figures 2 and 3, respectively. The field values were calculated as the scalar product (stdev*coeff) of the B-coefficient and the standard deviation of each variable. The

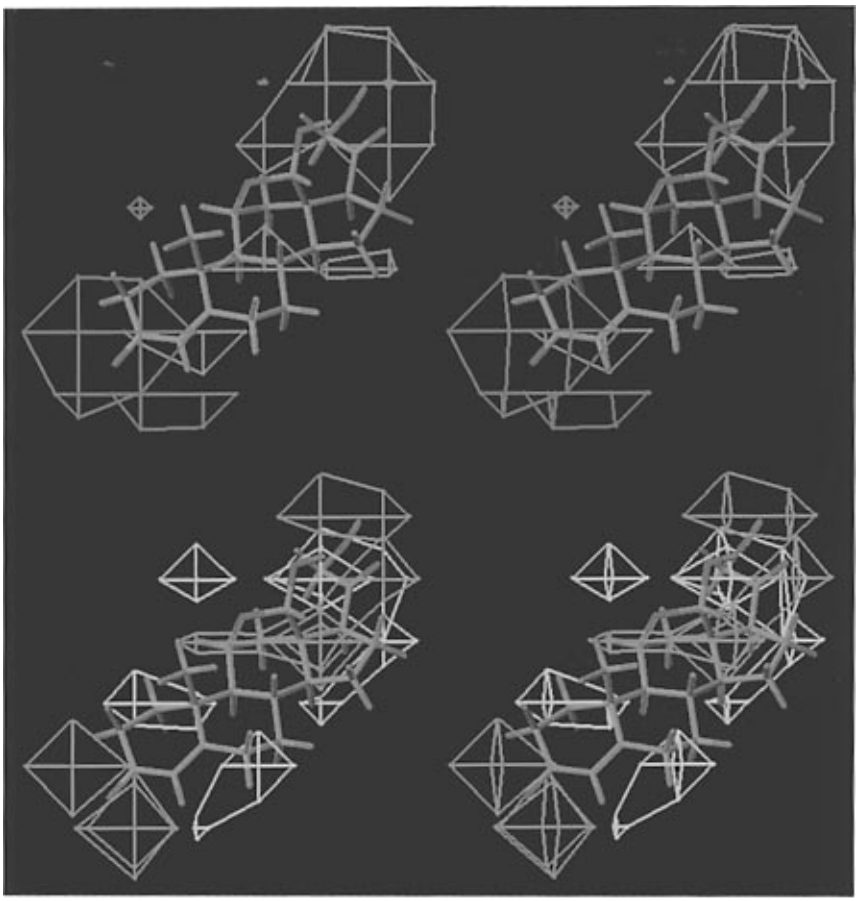

**Figure 2.** Steroview of the contour plot of steric field derived from PLS analysis. The contours were drawn at the +0.03 (magenta) level and −0.03 (cyan) level. (bottom) Stereoview of the contour plot of electrostatic field derived from PLS analysis. The contours were drawn at the +0.04 (green) level and −0.04 (yellow) level. Aldosterone is shown as the reference molecule.

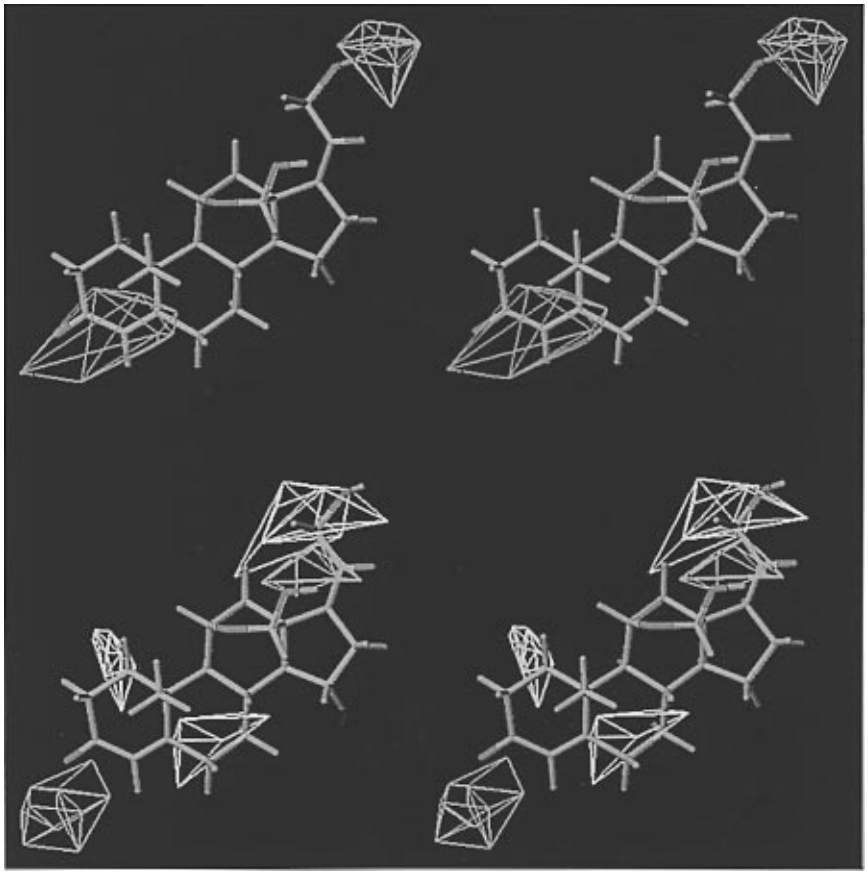

**Figure 3.** (top) Stereoview of the contour plot of steric field derived from PWPLS analysis. The contours were drawn at the +0.3 (magenta) level and −0.3 (cyan) level. (bottom) Stereoview of the contour plot of electrostatic field derived from PWPLS analysis. The contours were drawn at the +0.15 (green) level and −0.15 (yellow) level. Aldosterone is shown as the reference molecule.

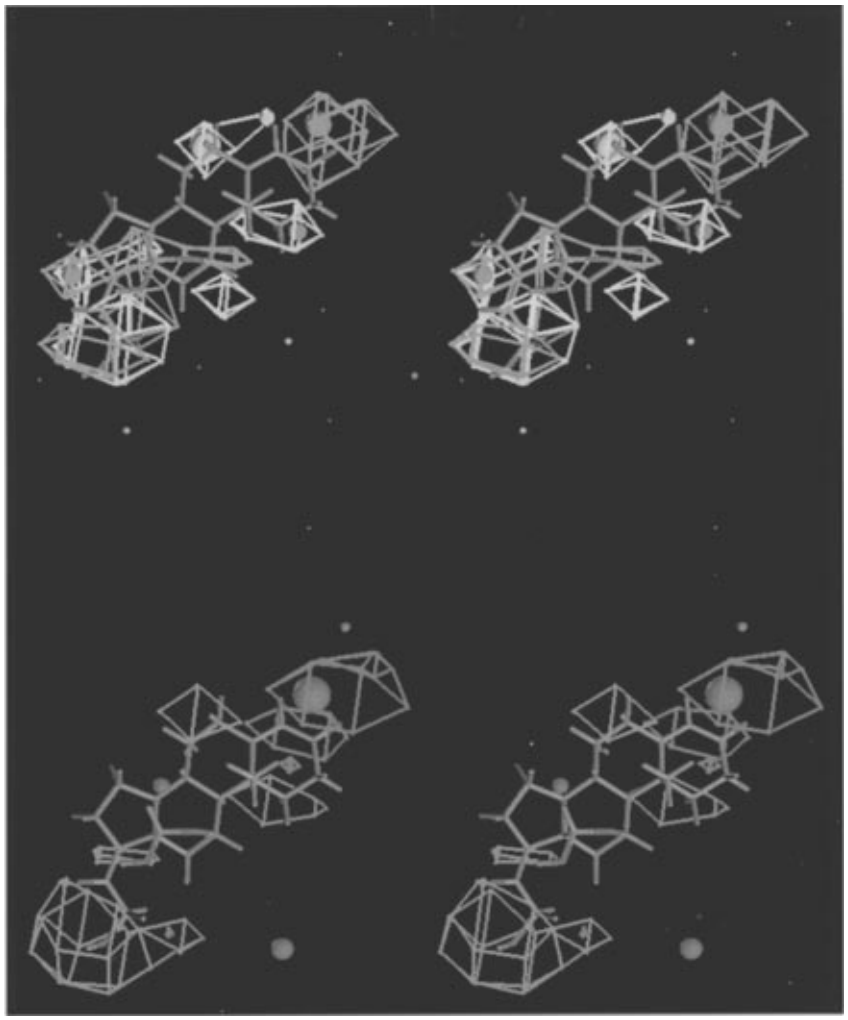

**Figure 4.** (top) Stereoview of the selected grid points of steric field by QOVS analysis and the contour plot derived from PLS analysis. The magenta grids are sterically favorable, while cyan grids are sterically unfavorable. (bottom) Stereoview of the selected grid points of electrostatic field by QOVS analysis and the contour plot derived from PLS analysis. The green grids are positive charge favorable, while yellow grids are negative charge favorable. Aldosterone is shown as the reference molecule.

contours display 3-D space areas, where small changes in molecular fields are strongly associated with changes of receptor-binding affinities. The interaction magnitudes were color-coded. In case of steric field, magenta contours surrounded regions where higher steric interference increase the binding affinity, and cyan contours indicate undesirable steric interaction where a decrease in steric interference improve the binding affinity. In case of electrostatic field, green contours surrounded regions where positively charged interference increase the binding affinity, and yellow contours indicating negatively charged interference improve the binding affinity. The contour maps of PLS steric and electrostatic fields with those of PWPLS steric and electrostatic fields resemble one another. $Q^2$s of PWPLS for both fields, however, were higher than that of PLS. Thus, in PWPLS, the essential grid points for CoMFA analysis were selected, and the noise grid points were omitted.

In QOVS, an insufficient number of grid points were selected to display contour plots. The results were visualized in Figure 4 as scalar products (stdev*coeff) so each selected grid point is displayed with different radii. The radius of each grid point was determined by stdev*coeff Å. Figure 4 also shows the contour maps derived from normal PLS. In steric field, the magenta grids are sterically favorable, while cyan grids are sterically unfavorable. In electrostatic field, the green grids are positive charge favorable, while the yellow grids are negative charge favorable. Some grids of both steric and electrostatic fields were included the contours of normal PLS. On the other hand, several contours had no grids. Thus QOVS models are substantially different from normal PLS models.

## CONCLUSION AND NOTES

We applied PWPLS and QOVS to the CoMFA analysis of CBG binding and could get more predictable models than that of normal PLS. As for steric field, $Q^2$s were improved form 0.719 (normal PLS) to 0.792 (PWPLS) and 0.875 (QOVS). As for electrostatic field, $Q^2$s were improved form 0.764 (normal PLS) to 0.773 (PWPLS) and 0.807 (QOVS). Furthermore, we could successfully visualize these predictable models.

## REFERENCES AND NOTES

(1) Cramer III, R. D.; 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.

(2) Raghavan, K.; Buolamwini, J. K.; Fesen, M. R.; Pommier, Y.; Kohn, K. W.; Weinstein, J. N. Three-Dimensional Quantitative Structure-Activity Relationship (QSAR) of HIV Integrase Inhibitors: A Comparative Molecular Field Analysis (CoMFA) Study. *J. Med. Chem.* **1995**, *38*, 890−897.

(3) Tsventan, G.; Gantchev, H. A.; Johan, E. L. Quantitative Structure-Activity Relationships/Comparative Molecular Field Analysis (QSAR/CoMFA) for Receptor-Binding Properties of Halogenated Estradiol Derivatives. *J. Med. Chem.* **1994**, *37*, 4164−4176.
(4) Waller, C. L.; Oprea, T. I.; Giolitti, A.; Marshall, G. R. Three-Dimensional QSAR of Human Immunodeficiency Virus (I) Protease Inhibitors. 1. A CoMFA Study Employing Experimentally-Determined Alignment Rules. *J. Med. Chem.* **1993**, *36*, 4152−4160.
(5) Waller, C. L.; Marshall, G. R. Three-Dimensional Quantitative Structure-Activity Relationship of Angiotensin-Converting Enzyme and the Rmolysin Inhibitors. II. A Comparison of CoMFA Models Incorporating Molecular Orbital Fields and Desolvation Free Energies Based on Active-Analog and Complementary-Receptor-Field Alignment Rules. *J. Med. Chem.* **1993**, *36*, 2390−2403.
(6) DePriest, S. A.; Mayer, D.; Naylor, C. B.; Marshall, G. R. 3D-QSAR of Angiotensin-Converting Enzyme and Thermolysin Inhibitors: a Comparison of CoMFA Models Based on Deduced and Experimentally Determined Active Site Geometries. *J. Am. Chem. Soc.* **1993**, *115*, 5372−5384.
(7) Atul, A.; Philip, P. P.; Ethan, W. T.; Hong, B. L.; Torsten, D.; Margareta, H.; Youhua, Y.; Georgina, L.; David, L. N.; John, W. R.; Arnold, R. M. Three-Dimensional Quantitative Structure-Activity Relationships of 5-HT Receptor Binding Data for Tetrahydropyridinylindole Derivatives: A Comparison of the Hansch and CoMFA Methods. *J. Med. Chem.* **1993**, *36*, 4006−4014.
(8) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. III. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735−743.
(9) Geladi, P.; Kowalski, B. R. Partial Least-Squares Regression: A Tutorial. *Anal. Chem.* **1986**, *185*, 1−17.
(10) Glem, W. G.; Dunn, W. J. III.; Scott, D. R. Principal Components Analysis and Partial Least Squares Regression. *Tetrahedron Computer Methodology* **1989**, *6*, 349−376.
(11) Baroni, M.; Clementi, S.; Cruciani, G.; Costantino, G.; Riganelli, D. Predictive Ability of Regression Models. Part II: Selection of the Best Predictive PLS Model. *J. Chemom.* **1992**, *6*, 347−356.
(12) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. *Quant. Struct. -Act. Relat.* **1993**, *12*, 9−20.
(13) Bode, W.; Wei, A. Z.; Huber, R.; Meyer, E.; Travis, J.; Neumann, S. X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *The EMBO Journal* **1986**, 5, *10*, 2453−2458.
(14) Tominaga, Y.; Fujiwara, I. Prediction-Weighted Partial Least-Squares Regression Method (PWPLS). *Chemom. Intell. Lab. Syst.* In press.
(15) Dunn, J. F.; Nisula, B. C.; Rodbard, D. Transport of Steroid Hormones: Binding of 21 Endogenous Steroids to Both Testosterone-Binding Globulin and Corticosteroid-Binding Globulin in Human Plasma. *J. Clin. Endocrinol. Metab.* **1981**, *53*, 58−68.
(16) Cramer III, R. D. Partial Least Squares (PLS): Its strengths and limitations. *Perspectives in Drug Discovery and Design* **1993**, *1*, 269−278.
(17) Sybyl Molecular Modeling Software, version 6.3; Tripos Associates, Inc.: St. Louis, MO 63144.
(18) Cruciani, G.; Watson, K. A. Comparative Molecular Field Analysis Using GRID Force-Field and GOLPE Variable Selection Methods in a Study of Inhibitor of Glycogen Phosphorylase b. *J. Med. Chem.* **1994**, *37*, 2589−2601.

CI970025Q