Novel 3D Descriptors Using Excluded Volume: Application to 3D Quantitative Structure−Activity Relationships

Yukio Tominaga* and Iwao Fujiwara

Department of Chemistry I, Discovery Research Laboratories I, Dainippon Pharmaceutical Company, Ltd., Enoki 33-94, Suita, Osaka 564, Japan

Received June 2, 1997[⊗]

Comparative molecular field analysis (CoMFA) has been used in drug design and three-dimensional quantitative structure−activity relationships (3D-QSAR). The success of a CoMFA study depends on the quality of the alignment rule. In other words, the alignment rule is the key in CoMFA analysis, while it is not a simple task to determine the rule. We developed novel 3D descriptors, the van der Waals excluded volume of each molecule and probe, which do not require alignment rules and are not significantly affected by the orientation of the molecules. Each probe is constructed by the excluded volume of two spheres with different radii and the center identical to each molecule's center of gravity. We applied the descriptors to QSAR analysis of the data set for the binding of 21 steroids to corticosteroid-binding globulin (CBG) which was introduced by Dunne et al. and compared the result with that of CoMFA.

INTRODUCTION

Since the study of Cramer et al.[1] was published, comparative molecular field analysis (CoMFA) has been successfully used in drug design and three-dimensional quantitative structure−activity relationships (3D-QSAR).[2−7] CoMFA provides a large number of variables which are generated to describe the nonbonded steric and/or electrostatic interaction energies between a number of probes and each molecule and uses partial least squares (PLS)[8−10] to correlate the changes in the observed biological activity with the changes in the chemical structure of a series of molecules.

The success of CoMFA is completely determined by the quality of the alignment rule, in other words, positioning of a molecular model within a fixed lattice. The alignment rule is the key input for CoMFA analysis, while the rule cannot be simply determined. To avoid this difficulty, we propose novel 3D descriptors.

We developed a probe constructed by the excluded volume of two spheres with different radii and the identical center corresponding to each molecule's center of gravity (Figure 1) and then descriptors which are the van der Waals excluded volume of each molecule and probe. A probe is layered like an onion. The descriptors are not significantly affected by the orientation of molecules. In CoMFA, on the other hand, if the orientation of the molecules is changed, interaction energies between molecules and each lattice point are changed accordingly (Figure 1). The descriptors represent the expansion of molecular volume in 3D space.

In addition to the descriptors derived from the entire molecule, we developed descriptors by determining the excluded volume between a specific type of atom of the molecule and probe. We classified an atom into 15 types. The descriptors derived from this method represent the expansion of a specific type of atom in 3D space.

In this paper, we applied these descriptors to QSAR analysis of the data set for the binding of 21 steroids to corticosteroid-binding globulin (CBG), which was introduced by Dunne et al.,[11] and compared the results with that of CoMFA.

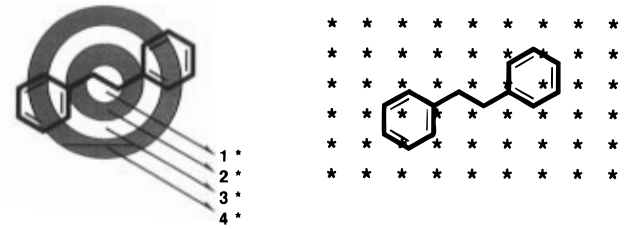

Figure 1. Schematic representation of the probe in this study (left) and CoMFA (right): ∗, layer nos.

METHOD

1. Data Set. The data set reported by Dunne et al. on the binding of 21 steroids to CBG was used.[11] Each 3D structure and the partial charge of each atom were taken directly from the SYBYL demonstration files, which are described in Cramer et al.[1]

2. System Used for Data Analysis. All calculations were carried out within SYBYL 6.3[12] on Indigo 2, running version 6.2 of the IRIX operating system.

3. Descriptors. The first probe (layer no. 1 in Figure 1) is an atom, e.g. iodide atom, I (van der Waals radius is 2.05 Å),[1] and the center of the probe is each molecule's center of gravity. The second probe (layer no. 2 in Figure 1) is constructed by 60 atoms (I); it means that the sphere is approximated by a polyhedron with 20 hexagonal (6 angled) surfaces and 12 pentagonal (5 angled) surfaces, which make the sphere's surface like a fullerene (C60). The first and second probes share the same center which corresponds to each molecule's center of gravity. The distance between the centers of the atom in the first probe and an atom in the second probe is 0.5 Å. The subsequent probes were also defined by using the method above. Descriptors were calculated by the MVOLUM command within SYBYL as the excluded volume between each molecule and probe. The notation of the obtained descriptors is represented as EV_{whole}.

[⊗] Abstract published in *Advance ACS Abstracts,* October 1, 1997.

S0095-2338(97)00230-8 CCC: $14.00

Table 1. CoMFA Analysis for 21 Steroids to Corticosteriod-Binding Globulin[a]

	Q^2
RUN1	0.156(1)
RUN2	−0.322(1)
RUN3	0.171(3)
RUN4	−0.207(1)
RUN5	0.112(1)
RUN6	−0.427(3)
RUN7	−0.083(3)
RUN8	0.080(4)
RUN9	−0.046(2)
RUN10	0.049(2)
MEAN	−0.052
SD[b]	0.207

[a] The table entries report the values of Q^2 and the number of components in parentheses. [b] Standard deviation.

The atoms of each molecule were classified into 15 types specified in SYBYL atom type notation:[13] C. Ar, C. 2, C. 3, all other types of carbon, O. 2, O. 3, all other types of oxygen, N. 2, N. 3, N. Ar, all other types of nitrogen, phosphorus, sulfur, halogen, and all other types of atoms except hydrogen. Each type of atom remained for each molecule, while other atoms were removed. The excluded volume between the remaining atoms and each probe were then evaluated. This procedure was repeated for all atom types. The notation of these descriptors is represented as EV_{type}. In this article, we used 21 layers, so EV_{whole}, EV_{type}, and EV_{both} consist of 21, 315, and 336 descriptors, respectively.

4. PLS Analysis. PLS analysis was carried out within SYBYL. All descriptors and biological activities were mean centered. Scaling in NONE. The most predictable model was determined by Q^2 (multiple correlation coefficient of predictions) from cross-validation with the leave-one-out method.

5. CoMFA Analysis. CoMFA analysis was carried out using the QSAR module of SYBYL. The aligned compounds were placed in a grid box containing equally spaced lattice points (2 Å). Grid box dimensions for CoMFA analysis were determined using the "create automatically" feature of the SYBYL/CoMFA program. The probe atom is Csp3. Non-covalent interaction energies (steric and electrostatic) were calculated for aligned compounds. The correlation of the non-covalent interaction energies with the target property was then analyzed using PLS. All CoMFA parameters and biological activities were mean centered. CoMFAstd was used for scaling. The most predictable model was derived from cross-validation with the leave-one-out method. All calculations were used option MINIMUM_SIGMA = 2.0.

RESULTS AND DISCUSSION

Using SPL within SYBYL, each molecule in the 21 QSAR data sets was randomly rotated with each molecule's center of gravity, and we made 10 randomly positioned data sets (RUNs 1−10).

CoMFA analysis was carried out for these randomly positioned data sets using steric and electrostatic fields. The results are summarized in Table 1. The Q^2 for RUNs 1−10 showed low values (from −0.427 to +0.156), indicating the molecules were not aligned. If the alignment was accurate, the Q^2 would have been 0.662, according to Cramer et al.[1]

Table 2. PLS Analysis for 21 Steroids to Corticosteroid-Binding Globulin Using EV_{whole}, EV_{type}, and EV_{both}.[a] The table entries report the values of Q^2 and the number of components in parentheses

	EV_{whole}	EV_{type}	EV_{both}
RUN	0.509(3)	0.856(7)	0.809(6)
RUN	0.576(3)	0.674(3)	0.659(3)
RUN3	0.267(1)	0.731(6)	0.624(3)
RUN4	0.263(1)	0.742(7)	0.622(4)
RUN5	0.289(1)	0.803(7)	0.641(3)
RUN6	0.492(3)	0.658(3)	0.640(3)
RUN7	0.443(4)	0.737(6)	0.620(3)
RUN8	0.258(1)	0.838(7)	0.761(4)
RUN9	0.306(1)	0.820(7)	0.781(7)
RUN10	0.279(1)	0.796(9)	0.685(4)
MEAN	0.368	0.766	0.694
SD[b]	0.123	0.068	0.072

[a] The table entries report the values of Q^2 and the number of components in parentheses. [b] Standard deviation.

For these 10 randomly positioned data sets, PLS analyses using EV_{whole}, EV_{type}, and EV_{both} were carried out. The results are shown in Table 2. The corresponding Q^2s differed in each RUN for all three descriptors because the descriptors were affected by the orientation of the molecules. Among the three descriptors, the mean of Q^2 for EV_{type} showed the highest Q^2 (0.766), while Q^2 of EV_{whole} was 0.368 and EV_{both} was 0.694. The standard deviation of EV_{type} was 0.068. The combination of the high Q^2 of 0.766 and the low standard deviation represented by PLS distinguishes the sound of the descriptors from their noises when the descriptors are correlated with the biological data. Although the molecules were not aligned precisely, the predictability ($Q^2 = 0.766 \pm 0.068$) of the 10 models consisting of EV_{type} was superior to that of CoMFA ($Q^2 = 0.662$), even when the molecules were aligned precisely.

We directed our attention to the standard deviation of Q^2 higher than 0.0 and estimated the sound and noise in the descriptors. Firstly, the difference of the descriptors from the viewpoint of the orientation of each molecule, noise, was examined for the 10 data sets. The standard deviations of the descriptors of RUNs 1−10 were evaluated for all 21 molecules. The mean of the standard deviations of the descriptors derived from 21 molecules was then evaluated. The mean values are the noise of the descriptors and are summarized in Table 3. To measure the sound of the same data sets, the standard deviations of the descriptors derived from different molecules in each RUN were evaluated. The mean of the standard deviations of the descriptors was then evaluated for all RUNs. The mean values are the sounds of the descriptors, and are summarized in Table 4. The ratio of sound to noise (S/N) varied from 1 to 115. However, the only time S/N becomes low is when both sound and noise are small. This indicates that, in our descriptors, the noise is small enough to construct the QSAR model with PLS.

We further examined the potential of the atoms other than iodide as a probe. Iodide atoms (I) with a van der Waals radius of 2.05 Å were replaced with carbon atoms (C) with a van der Waals radius of 1.52 Å or hydrogen atoms (H) with a van der Waals radius of 1.08 Å.[1] The descriptors, EV_{whole} and EV_{type}, were evaluated using the probes constructed with carbon atoms or hydrogen atoms, and then PLS analysis was carried out. The results are shown in Table 5 for C probes and Table 6 for H probes. In both cases of C probes and H probes, the mean of Q^2 for EV_{type} showed the

Table 3. Noise of Descriptors

layer no.	EV_{whole}	EV_{type} (O. 3)	EV_{type} (O. 2)	EV_{type} (C. Ar)	EV_{type} (C. 2)	EV_{type} (C. 3)
1	0.16	0.02	0.00	0.02	0.06	0.19
2	0.25	0.03	0.01	0.03	0.10	0.31
3	0.37	0.06	0.01	0.04	0.18	0.45
4	0.47	0.07	0.01	0.04	0.23	0.53
5	0.64	0.14	0.04	0.05	0.29	0.66
6	0.67	0.15	0.11	0.05	0.33	0.69
7	0.78	0.22	0.22	0.05	0.37	0.76
8	0.84	0.36	0.35	0.07	0.44	0.83
9	0.95	0.48	0.53	0.09	0.52	0.89
10	1.19	0.58	0.68	0.09	0.51	0.80
11	1.33	0.71	0.69	0.08	0.47	0.79
12	1.35	0.78	0.72	0.07	0.48	0.76
13	1.68	0.89	0.89	0.07	0.65	0.93
14	1.89	1.06	1.10	0.12	0.76	1.09
15	2.01	1.23	1.30	0.13	0.74	1.10
16	2.00	1.17	1.31	0.10	0.52	0.71
17	1.69	0.89	0.97	0.03	0.24	0.21
18	1.03	0.51	0.51	0.00	0.05	0.01
19	0.44	0.17	0.22	0.00	0.00	0.00
20	0.10	0.01	0.04	0.00	0.00	0.00
21	0.00	0.00	0.00	0.00	0.00	0.00

Table 4. Sounds of Descriptors

layer no.	EV_{whole}	EV_{type} (O. 3)	EV_{type} (O. 2)	EV_{type} (C. Ar)	EV_{type} (C. 2)	EV_{type} (C. 3)
1	0.33	1.29	0.33	1.73	2.32	0.79
2	1.18	2.39	0.87	3.45	3.90	2.28
3	2.62	3.64	1.56	5.73	5.68	4.43
4	4.48	4.60	2.11	8.15	7.45	6.96
5	6.96	5.36	2.35	10.75	9.32	10.05
6	9.46	5.95	2.52	13.36	11.15	13.12
7	11.72	6.32	3.05	15.84	13.17	15.86
8	13.87	6.50	3.86	17.77	15.42	18.24
9	15.85	6.78	4.87	18.28	17.13	19.62
10	17.21	6.99	5.91	17.18	17.55	19.36
11	17.97	7.04	6.77	14.71	16.90	17.58
12	18.25	6.97	7.46	11.78	15.20	15.04
13	16.56	6.45	7.80	8.76	12.55	11.85
14	14.04	5.46	7.37	5.64	9.06	8.16
15	11.22	4.39	6.36	2.90	5.45	4.97
16	8.27	3.34	4.85	0.91	2.40	2.32
17	5.31	2.27	3.06	0.08	0.73	0.56
18	2.76	1.14	1.53	0.00	0.10	0.03
19	1.03	0.32	0.58	0.00	0.00	0.00
20	0.17	0.01	0.08	0.00	0.00	0.00
21	0.00	0.00	0.00	0.00	0.00	0.00

Table 5. PLS Analysis for 21 Steroids to Corticosteroid-Binding Globulin using EV_{whole}, EV_{type}, and EV_{both}[a]

	EV_{whole}	EV_{type}	EV_{both}
RUN1	0.559(6)	0.854(8)	0.842(6)
RUN2	0.423(3)	0.665(3)	0.642(3)
RUN3	0.235(1)	0.608(3)	0.588(3)
RUN4	0.224(1)	0.695(6)	0.535(2)
RUN5	0.281(3)	0.788(8)	0.625(3)
RUN6	0.451(3)	0.609(3)	0.604(3)
RUN7	0.314(1)	0.608(3)	0.577(3)
RUN8	0.455(3)	0.757(3)	0.755(5)
RUN9	0.328(1)	0.787(6)	0.751(7)
RUN10	0.281(1)	0.662(3)	0.635(3)
MEAN	0.355	0.703	0.655
SD[b]	0.111	0.088	0.096

[a] The table entries report the values of Q^2 and the number of components in parentheses. [b] Standard deviation.

highest of the three descriptors: 0.703 for C probes and 0.658 for H probes. These values were inferior to that of I probes ($Q^2 = 0.766$). The standard deviations of Q^2 for the C probe (0.088) and the H probe (0.100) were larger than that of the I probe (0.068). I probes showed the best result among the three probes, with the highest mean Q^2 and the lowest standard deviation of Q^2.

Table 6. PLS Analysis for 21 Steroids to Corticosteroid-Binding Globulin using EV_{whole}, EV_{type}, and EV_{both}[a]

	EV_{whole}	EV_{type}	EV_{both}
RUN1	0.257(1)	0.801(7)	0.692(3)
RUN2	0.260(1)	0.634(3)	0.598(2)
RUN3	0.185(1)	0.540(2)	0.522(2)
RUN4	0.183(1)	0.666(6)	0.499(2)
RUN5	0.187(1)	0.785(2)	0.583(2)
RUN6	0.480(3)	0.551(2)	0.569(2)
RUN7	0.225(1)	0.520(2)	0.490(2)
RUN8	0.629(6)	0.729(3)	0.685(3)
RUN9	0.296(1)	0.715(6)	0.620(2)
RUN10	0.263(1)	0.638(3)	0.612(2)
MEAN	0.296	0.658	0.587
SD[b]	0.146	0.100	0.070

[a] The table entries report the values of Q^2 and the number of components in parentheses. [b] Standard deviation.

Table 7. PLS Analysis for 21 Steroids for Corticosteroid-Binding Globulin using EV_{whole}, EV_{type}, and EV_{both}[a]

	EV_{whole}	EV_{type}	EV_{both}
RUN1	0.411(3)	0.762(6)	0.772(6)
RUN2	0.519(6)	0.381(3)	0.276(1)
RUN3	0.480(3)	0.625(6)	0.629(6)
RUN4	0.219(1)	0.595(7)	0.521(9)
RUN5	0.412(3)	0.407(2)	0.399(2)
RUN6	0.259(1)	0.659(6)	0.633(6)
RUN7	0.232(1)	0.315(3)	0.349(3)
RUN8	0.253(1)	0.353(2)	0.334(2)
RUN9	0.293(1)	0.293(1)	0.351(1)
RUN10	0.348(1)	0.453(6)	0.571(7)
MEAN	0.343	0.484	0.484
SD[b]	0.108	0.163	0.165

[a] The table entries report the values of Q^2 and the number of components in parentheses. [b] Standard deviation.

We then changed the distance between atoms of each probe layer from 0.5 Å to 1.0 Å. Accordingly, the number of layers decreased from 21 to 11. The number of descriptors were also decreased from 21 to 11 for EV_{whole}, from 315 to 165 for EV_{type}, and from 336 to 176 for EV_{both}. These descriptors using I atoms were evaluated, and then PLS analysis was carried out. The results are summarized in Table 7. The means of Q^2 were 0.484 for both EV_{type} and EV_{both} and 0.343 for EV_{whole}. These values were inferior to the mean where the distance is 0.5 Å. The standard deviations were 0.163 for EV_{type}, 0.165 for EV_{both}, and 0.108 for EV_{whole}, which are larger than the standard deviations where the distance is 0.5 Å. This indicates the distance of 0.5 Å between the atoms of each probe is more preferable than 1.0 Å.

Through the 3D-QSAR of the CBG case, the usefulness of the descriptors was evaluated. The predictability of the models ($Q^2 = 0.766 \pm 0.068$) were superior to that of the well-aligned CoMFA model ($Q^2 = 0.662$). In the CBG case, it is not necessary to select the conformation for each compound because each compound is a rigid steroid. The selection of conformation, however, is as important in a CoMFA study as the alignment rule. Further validation of the descriptors about the conformation problem is the subject of ongoing research.

CONCLUSIONS

The 3D descriptors we developed derive predictable 3D-QSAR models through a simple procedure which does not require alignment rules. The predictability of the models ($Q^2 = 0.766 \pm 0.068$) were even superior to that of the well-aligned CoMFA model ($Q^2 = 0.662$).

REFERENCES AND NOTES

(1) Cramer, R. D., III; Patternson, 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 Comprision 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 Comput. Methodol.* **1989**, *6*, 349−376.

(11) Dunn, J. F.; Nisula, B. C.; Rodbard, D. Transport of Steroid Hormones: Binding of 21 Endogeneous Steroids to Both Testosterone-Binding Globulin and Corticosteroid-Binding Globulin in Human Plasma. *J. Clin. Endocrinol. Metab.* **1981**, *53*, 58−68.

(12) *Sybyl Molecular Modeling Software*, Version 6.3; Tripos Associates, Inc.: St. Louis, MO.

(13) Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, F. R. D.; Weinberger, L. E. Neighborhood Behavior: A Useful Concept for Validation of "Molecular Diversity" Descriptors. *J. Med. Chem.* **1996**, *39*, 3049−3059.

CI9702306