Journal of Computer-Aided Molecular Design, 7 (1993) 263–280
ESCOM

J-CAMD 200

Use of the hydrogen bond potential function in a comparative molecular field analysis (CoMFA) on a set of benzodiazepines

Ki Hwan Kim[a,*], Giovanni Greco[b], Ettore Novellino[b], Carlo Silipo[b] and Antonio Vittoria[b]

[a]*Computer Assisted Molecular Design, Pharmaceutical Products Division, Abbott Laboratories, Abbott Park, IL 60064, U.S.A.*
[b]*Dipartimento di Chimica Farmaceutica e Tossicologica, Università di Napoli, 80131 Naples, Italy*

Received 2 June 1992
Accepted 18 November 1992

Key words: 3D-QSAR; CoMFA; GRID; Hydrogen-bond potential; Partial Least Squares; Hydrophobicity; Benzodiazepine

SUMMARY

The results of the GRID-Comparative Molecular Field Analysis (CoMFA) were compared with those of the SYBYL-CoMFA in a study of benzodiazepines. The results demonstrate that the hydrogen bonding function using the GRID H_2O probe in a CoMFA can successfully describe the hydrophobic effects of substituents without any bias or preconcept of their effects in the development.

INTRODUCTION

A relatively recent development in the area of structure–activity relationships [1] is represented by Comparative Molecular Field Analysis (CoMFA) [2]. In this 3D-QSAR approach a number of steric and electrostatic energy values is sampled at regularly spaced points of a lattice in which the molecules of a data set have been superimposed. These energies are measured as interactions between each ligand and properly selected probe through simple functions of molecular mechanics force fields. The intermolecular energies obtained are then correlated with biological activity using the partial least squares (PLS) method [3–5]. The most predictive model is chosen from cross-validation procedures [6,7]. CoMFA has been employed not only in pharmacodynamic studies [8–12] but also to model chemical equilibrium constants and traditional QSAR descriptors [13–15]. Kim [16,17] has observed that the GRID [18–20] H_2O probe can be an efficient hydrophobic descriptor in CoMFA.

*To whom correspondence should be addressed.

Recently, Greco et al. [12] reported a study of a combined QSAR-CoMFA approach on a set of agonistic benzodiazepines with the SYBYL force field [21]. In the QSAR analysis the following parameters were employed as molecular descriptors in the final equation (Eq. 1):

$$pIC_{50} = 0.991(\pm 0.255)I_{2'} - 0.035(\pm 0.014)E_{HOMO} + 0.409(\pm 0.180)\pi_7 - 0.407(\pm 2.990) \quad (1)$$
$$n = 48,\ s = 0.410,\ r = 0.867,\ F_{1,44} = 20.6$$

In Eq. 1, the substituent constant [22], π_7, describes the hydrophobicity of the substituents on the benzodiazepine ring at 7-position corresponding to the 7-chlorine atom in diazepam (**I**), E_{HOMO} is the AM1 calculated HOMO energy [23], and $I_{2'}$ is an indicator variable which held the value of 1 when the 2'-position of the C ring was substituted. The positive coefficient of π_7 suggested that the substitutent at the 7-position may interact with a hydrophobic pocket of the receptor. Additionally, charge–transfer interactions with an electronegative site appear to promote the binding affinity. The actual role played by the 2'-substituents (mostly halogens) in increasing the binding affinity could not be properly assessed because of the poor structural variance at the 2'-position. The molecular modeling studies, however, excluded that the 2'-substituent could influence the strength of the binding through intramolecular conformational effects [12].

Since the results of the traditional QSAR provided evidence of an involvement of a specific ligand–receptor hydrophobic interaction, an attempt was made to properly include the π_7 parameter in the CoMFA model. On the basis of empirical observations and theoretical considerations [12], the π_7 descriptor was added, through a weighting factor of 100, to a CoMFA data table whose columns were constituted by electrostatic energies. The weighted π_7 substituent constant plus the CoMFA electrostatic field explained more than 75% of the variance in the biological data.

Although the introduction of a hydrophobic parameter in a CoMFA calculation was successfully accomplished, there are some unavoidable limitations in such an approach. First, a set of π or log P values should be available for all the investigated compounds. Second, one or more substituent positions must be identified for their supposed hydrophobic relevance on a subjective basis.

In light of our experiences with the hydrogen bond potential function in describing hydrophobic interactions [16,17], the set of benzodiazepines was evaluated for their binding affinity potency with the hydrogen bond function of GRID in this CoMFA study. The results are compared with the SYBYL-CoMFA results obtained previously [12].

METHODS

Molecular modeling

The geometries of the compounds were modeled with the standard bond distances and angles using SYBYL and optimized with the AM1 [23] of MOPAC [24]. All the molecules were considered exclusively in the conformation in which the C3 carbon atom is positioned above the plane of the aromatic A ring. The molecules were superimposed on diazepam (**I**) by minimizing the root mean square of the interatomic distances using the C2 carbonyl oxygen or an 'equivalent' nitrogen atom in that position, the N4 nitrogen or an 'equivalent' carbon atom in that position, and

the centroids of A and C rings. Detailed molecular modeling procedures are described in a previous publication [12].

CoMFA interaction energy calculation

The 'hydrophobic', steric, and electrostatic potential energy fields of each molecule were calculated at various lattice points surrounding the molecule using H_2O, CH_3, and H^+ probes with the program GRID as described in a previous communication [14,15,17]. A van der Waals radius of 1.70 and a charge of 0.0 were used for the H_2O probe with two hydrogen-donating and two hydrogen-accepting properties. The polarizability and effective number of electrons were 1.20 and 7, respectively. The optimum hydrogen-bond energy was −4.0 kcal/mol and the optimum hydrogen-bond length was 2.8 Å for O···O hydrogen bonds. These are currently the default values in GRID. In GRID, the H_2O probe was treated as if they can rotate freely and the length and orientational dependence of the hydrogen-bond were considered. A van der Waals radius of 1.95 and a charge of 0.0 were used for the CH_3 probe, and a zero van der Waals radius and a charge of 1.0 were used for the H^+ probe.

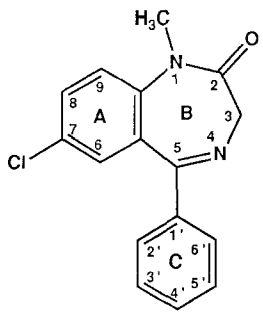
For each substituted benzodiazepine molecule, the energies at a total of 1573 grid points were calculated with 2 Å spacing in a lattice of $24 \times 20 \times 20$ (X = −14 to 10, Y = −9 to 11, Z = −9 to 11). All the energy values with a value greater than 4.0 kcal/mol were truncated to 4.0. Any lattice point for which the standard deviation of the energies was less than 0.05 was discarded. These procedures reduced the number of lattice points to 221, 239, and 1349 for the H_2O, CH_3 or H^+ probe, respectively.

Partial least squares (PLS) calculations

Ten to fifteen orthogonal latent variables were first extracted by the standard PLS algorithm. The number of latent variables extracted was always at least 3 less than the number of the compounds included in the correlation. These latent variables were subjected to the PLS validation test in the original order of extraction or in the order of their correlation with the dependent variable. The 'best' correlation model was chosen on the basis of which model significantly minimized the sum of squares of the difference in activity between the predicted and observed values using predictions made from a leave-one-out cross-validation test. After the number of latent variables was established from the cross-validation, the 'best' correlation model was derived with the same number of latent variables. The final model was further validated by the overall and the stepwise F-statistics. If F-statistics did not support the model, the least significant latent variable was eliminated and the model was re-derived. The variables $Z1_{H_2O}$–$Z3_{H_2O}$ and $Z1_{CH_3}$–$Z3_{CH_3}$ in the correlation equations are the first through the third latent variables from H_2O and CH_3 probes, respectively. The variables $Z1_{H^+}$–$Z3_{H^+}$ are the first through the third latent variables from an H^+ probe, and the variables $Z1_{H_2O,H^+}$–$Z7_{H_2O,H^+}$ are the first through the seventh latent variables from both H_2O and H^+ probes respectively obtained from the PLS analysis in each example. PRESS and press s are calculated as follows: $PRESS_{(n)} = \Sigma\{Y_{obs}-Y_{cal(n)}\}^2$ and press $s_{(n)} = \{PRESS_{(n)}/N\}^{1/2}$, where n is the number of latent variables, Y_{obs} and Y_{cal} are the observed and calculated biological activity values, and N is the number of compounds included in the correlations.

RESULTS AND DISCUSSION

In the GRID-CoMFA [25] study, a heterogeneous set of 48 benzodiazepine analogues was used. These compounds are listed in Table 1 and shown in Fig. 1. The structural modifications involve various positions of the seven-membered B-ring and the nature of A-ring substituent which corresponds to the 7-chlorine atom in diazepam (**I**).



(I)

The biological activity of interest is the ligand binding affinity (pIC_{50}) to the benzodiazepine receptors determined as competitive binding by the analogues to [^{3}H]diazepam. The measured binding affinity data were originally reported by Haefely et al. [26].

The geometries of the compounds were optimized with AM1 of MOPAC [24]. All the molecules were considered exclusively in the conformation in which the C3 carbon atom is positioned above the plane of the aromatic A ring. The molecules were superimposed on diazepam (**I**) using the C2 carbonyl oxygen or an 'equivalent' nitrogen atom in that position, the N4 nitrogen or an 'equivalent' carbon atom in that position, and the centroids of A and C rings. Details of the molecular modeling performed on these compounds are described in a previous study [12].

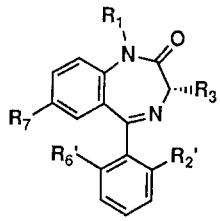
Three GRID probes were used in the interaction energy calculations; H_2O was used for hydrophobic (including hydrogen bonding), CH_3 for steric, and H^+ for electrostatic interactions.

The effects of grid lattice position on the correlation were first investigated. The grid box for the lattice points was shifted by −0.5, −1.0, and −1.5 Å in each of the directions X, Y, and Z. The statistics from the CoMFA results at these different lattice points are summarized in Table 2. In general, the results from different lattice positions are comparable as we had also observed in other cases [15]. Others sometimes observed significantly different results [2]. The correlations from the lattice points offset by either −1.0 or −1.5 Å gave better fitted and cross-validated results than those by −0.5 Å or the original grid position. The H_2O and H^+ probes also gave more consistent results than the CH_3 probe regardless of the locations of the grid box. Based on the results, the lattice position that was shifted from the original position by −1.0 Å was chosen as the best location of the lattice points and used for the following investigations in this study. In this lattice position, all three probes gave good results.

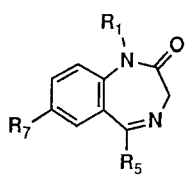
In the chosen lattice position, the H_2O probe gave the best correlation (Eq. 2), followed by the CH_3 and H^+ probes. There are high correlations between the latent variables Z_{H_2O}'s and Z_{CH_3}'s especially with the first two components with the set of compounds included in this study (see

TABLE 1
OBSERVED AND CALCULATED Log $1/IC_{50}$ VALUES USING EQ. 2

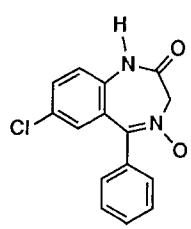
		Log $1/IC_{50}$			PLS latent variables		
No.	Compound	obs	calc	dev	$Z1_{H_2O}$	$Z2_{H_2O}$	$Z3_{H_2O}$
1	RO-054318	6.34	6.25	0.09	−7.891	−4.754	−2.619
2	RO-053072	6.41	6.39	0.02	−8.610	−0.030	−2.389
3	RO-054528	6.42	7.41	−0.99	−0.573	−0.872	−1.881
4	RO-052921	6.45	6.73	−0.28	−8.002	1.040	0.509
5	RO-207736	7.02	7.27	−0.25	−0.926	−1.289	−2.929
6	RO-054619	7.12	7.27	−0.15	−3.240	3.455	−1.893
7	RO-205397	7.37	7.39	−0.02	−1.417	2.844	−3.027
8	RO-053061	7.40	6.90	0.50	−7.012	1.641	0.712
9	RO-202533	7.44	7.56	−0.12	−3.494	4.156	2.064
10	RO-202541	7.52	8.03	−0.51	2.501	1.213	−0.089
11	RO-205747	7.62	7.74	−0.12	−3.421	5.131	3.758
12	RO-054336	7.68	7.22	0.46	−5.569	2.837	1.868
13	RO-203053	7.74	8.42	−0.68	1.419	7.581	3.107
14	RO-052904	7.89	7.91	−0.02	−1.408	4.912	2.653
15	Diazepam	8.09	7.51	0.58	−2.029	−1.325	2.254
16	RO-116896	8.15	8.23	−0.08	6.830	−0.946	−3.524
17	RO-075220	8.26	8.25	0.01	4.292	0.600	0.200
18	RO-143074	8.27	7.94	0.33	−2.794	6.660	4.387
19	Flunitrazepam	8.42	8.11	0.31	3.955	1.332	−1.703
20	RO-053590	8.45	8.50	−0.05	5.990	7.041	−3.490
21	RO-079957	8.54	8.32	0.22	1.954	2.480	3.949
22	RO-133780	8.62	8.28	0.33	1.866	1.901	4.030
23	RO-053367	8.70	8.23	0.47	0.490	6.195	2.974
24	Delorazepam	8.74	8.53	0.21	3.116	7.694	1.589
25	Clonazepam	8.74	8.40	0.20	5.028	6.935	−3.026
26	RO-0.54435	8.82	8.08	0.73	2.513	5.623	−2.189
27	Meclonazepam	8.92	8.75	0.17	8.741	5.609	−3.892
28	RO-053328	7.06	7.91	−0.85	−1.602	4.643	3.235
29	Tetrazepam	7.47	7.37	0.09	−3.016	−1.660	2.339
30	Bromazepam	7.74	7.24	0.50	−4.530	2.610	0.482
31	Demoxepam	6.51	6.88	−0.37	−3.859	0.108	−4.116
32	Chlordiazepoxide	6.45	6.39	0.06	−3.850	−9.255	−4.861
33	Alprazolam	7.70	7.41	0.29	0.142	−6.770	0.519
34	Triazolam	8.40	8.36	0.04	5.732	−3.031	1.378
35	RO-116679	8.40	8.05	0.35	7.484	−6.548	−3.683
36	RO-221892	7.92	8.18	−0.26	4.875	−9.658	4.722
37	RO-215205	8.13	8.38	−0.25	5.075	−8.476	6.317
38	Midazolam	8.32	8.32	0.00	4.119	−7.807	6.762
39	RO-141359	7.15	7.77	−0.62	0.693	2.735	−1.555
40	RO-158867	7.60	8.14	−0.54	5.946	3.974	−6.403
41	RO-147187	6.39	6.32	0.07	−5.105	−10.61	−2.754
42	RO-159270	8.30	7.99	0.31	7.631	−4.567	−5.981
43	Desmethylclobazam	6.68	6.84	−0.16	−3.610	−1.053	−4.262
44	Clobazam	6.89	6.69	0.20	−2.972	−6.187	−4.185
45	RO-221251	7.96	8.20	−0.24	3.620	−7.213	5.546
46	RO-223245	8.55	8.48	0.06	5.448	−4.823	4.771
47	RO-223148	6.38	6.32	0.06	−7.292	−6.585	−1.465
48	Premazepam	6.77	6.99	−0.22	−7.238	2.537	1.792



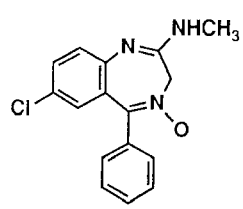
No.	R_7	R_1	R_3	$R_{2'}$	$R_{6'}$
1	NH_2	CH_3	H	H	H
2	NH_2	H	H	H	H
3	CN	CH_3	H	H	H
4	H	H	H	H	H
5	NHOH	CH_3	H	F	H
6	NH_2	H	H	Cl	H
7	CHO	H	H	H	H
8	F	H	H	H	H
9	Et	H	H	H	H
10	CN	CH_3	H	F	H
11	$CH=CH_2$	H	H	H	H
12	H	H	H	F	H
13	$COCH_3$	H	H	F	H
14	CF_3	H	H	H	H
15	Cl	CH_3	H	H	H
16	NO_2	CH_3	(S)-CH_3	F	H
17	Cl	CH_3	H	Cl	Cl
18	N_3	H	H	F	H
19	NO_2	CH_3	H	F	H
20	NO_2	H	H	CF_3	H
21	I	CH_3	H	F	H
22	Br	CH_3	H	F	F
23	Cl	H	H	F	H
24	Cl	H	H	Cl	H
25	NO_2	H	H	Cl	H
26	NO_2	H	H	F	H
27	NO_2	H	(S)-CH_3	Cl	H



No.	R_7	R_1	R_5
28	Cl	CH_3	
29	Cl	H	
30	Br	H	

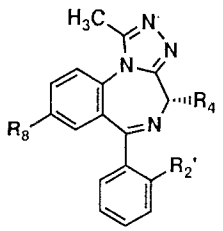


31

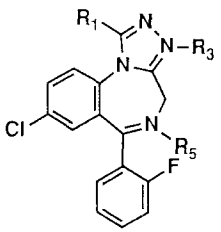


32

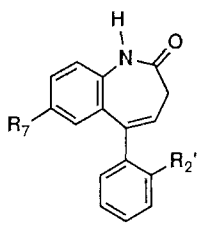
Fig. 1. Compounds used in this study.



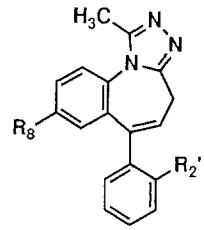
No.	R_8	R_4	$R_{2'}$
33	Cl	H	H
34	Cl	H	Cl
35	NO_2	(S)-CH_3	F



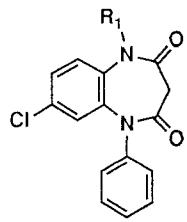
No.	R_1	R_3	R_5
36	H	COO-i-Pr	O
37	H	$COOCH_3$	
38	CH_3	H	



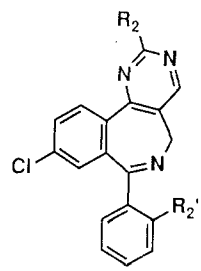
No.	R_7	$R_{2'}$
39	Cl	F
40	NO_2	Cl



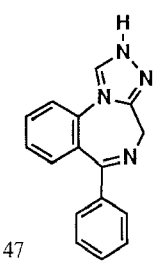
No.	R_8	$R_{2'}$
41	H	H
42	NO_2	Cl



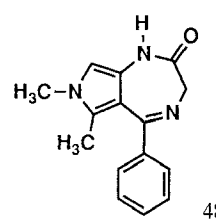
No.	R_1
43	H
44	CH_3



No.	R_1	$R_{2'}$
45	CH_3	H
46	H	Cl



47



48

Fig. 1. (continued)

TABLE 2
SUMMARY OF STATISTICS FROM MODELS AT DIFFERENT LATTICE POSITIONS

Lattice offset	H_2O probe				CH_3 probe				H^+ probe			
	L	s	r	press s	L	s	r	press s	L	s	r	press s
0.0Å	3	0.436	0.849	0.552	1	0.519	0.765	0.567	4	0.503	0.797	0.573
−0.5Å	3	0.411	0.866	0.523	3	0.468	0.823	0.616	4	0.509	0.791	0.588
−1.0Å	3	0.383	0.885	0.502	3	0.426	0.856	0.544	4	0.504	0.797	0.579
−1.5Å	3	0.395	0.878	0.483	2	0.472	0.815	0.562	4	0.510	0.790	0.587

L = Number of latent variables in the model; s = standard error of estimation; r = correlation coefficient of the fitted model; press s = standard error of estimation of the leave-one-out cross-validation test.

Table 3). For example, the correlation coefficient between $Z1_{H_2O}$ and $Z1_{CH_3}$ is 0.94. Therefore, it is not surprising to see a high correlation with the CH_3 probe, even though the quality is not as good as that of the H_2O probe. Likewise, there are significant correlations between the latent variables Z_{H_2O}'s and Z_{H^+}'s especially with the first component.

Equation 2 is the 'best' correlation model chosen from the H_2O probe. In this equation, pIC_{50} is the competitive binding affinity by the analogues of [^{3}H]diazepam to the benzodiazepine receptors. In the correlation equation, n is the number of compounds used, s is the residual standard deviation, r is the correlation coefficient and press s is the standard deviation from the leave-one-out cross-validation. F and P are the F-statistics and significance probability, respectively.

$$pIC_{50} = 0.127(\pm0.012)Z1_{H_2O} + 0.046(\pm0.011)Z2_{H_2O} + 0.073(\pm0.016)Z3_{H_2O} + 7.665(\pm0.055) \quad (2)$$
$$n = 48,\ s = 0.383,\ r = 0.885,\ F = 53.2,\ P = 0.0001,\ \text{press s} = 0.502$$

The hydrophobic effects (H_2O probe) alone could explain 78% of the variance in the data. The observed and calculated biological activity of the compounds using Eq. 2 and the latent variables used are given in Table 1.

Equations 3 and 4 show the effects of steric and electrostatic effects (CH_3 and H^+ probes). A rather significant correlation was obtained with the CH_3 probe because of the high collinearity between the variables from the H_2O and CH_3 probes with the present set of compounds. Inclusion of additional latent variables from the H^+ probe in the model did not significantly improve the correlation already obtained from the CH_3 probe as much as the correlation from the H_2O probe. There is also a certain amount of collinearity between the first latent variables of the H_2O model and the H^+ model (see Table 3) when they were derived independently. The correlation coefficient between $Z1_{H_2O}$ and $Z1_{H^+}$ is 0.69.

$$pIC_{50} = 0.147(\pm0.015)Z1_{CH_3} + 0.044(\pm0.015)Z2_{CH_3} + 0.083(\pm0.022)Z3_{CH_3} + 7.665(\pm0.062) \quad (3)$$
$$n = 48,\ s = 0.426,\ r = 0.856,\ F = 40.1,\ P = 0.0001,\ \text{press s} = 0.544$$

$$pIC_{50} = 0.076(\pm0.013)Z1_{H^+} + 0.083(\pm0.021)Z2_{H^+} + 0.238(\pm0.055)Z3_{H^+} + 0.094(\pm0.043)Z4_{H^+} + 7.665(\pm0.073) \quad (4)$$
$$n = 48,\ s = 0.504,\ r = 0.797,\ F = 18.7,\ P = 0.0001,\ \text{press s} = 0.579$$

TABLE 3
PEARSON CORRELATION COEFFICIENTS BETWEEN THE PLS LATENT VARIABLES

	$Z1_{H_2O}$	$Z2_{H_2O}$	$Z3_{H_2O}$	$Z1_{CH_3}$	$Z2_{CH_3}$	$Z3_{CH_3}$	$Z1_{H^+}$	$Z2_{H^+}$	$Z3_{H^+}$
$Z1_{H_2O}$	1.00	−0.00	−0.00	0.94	−0.07	0.10	0.69	0.18	0.46
$Z2_{H_2O}$		1.00	−0.00	0.07	0.93	−0.05	0.28	0.28	−0.28
$Z3_{H_2O}$			1.00	−0.01	0.02	0.56	−0.39	0.34	0.22
$Z1_{CH_3}$				1.00	−0.00	0.00	0.65	0.18	0.40
$Z2_{CH_3}$					1.00	0.00	0.25	0.31	−0.33
$Z3_{CH_3}$						1.00	−0.02	0.48	−0.01
$Z1_{H^+}$							1.00	−0.00	−0.00
$Z2_{H^+}$								1.00	0.00
$Z3_{H^+}$									1.00

Equation 5 was obtained from both the H_2O and H^+ probes. The addition of the electrostatic effects to the hydrophobic effects improved the correlation significantly explaining 96% of the total variance in the binding affinity.

$$\begin{aligned} pIC_{50} = {} & 0.084(\pm 0.004)Z1_{H_2O,H^+} + 0.079(\pm 0.006)Z2_{H_2O,H^+} + 0.040(\pm 0.006)Z3_{H_2O,H^+} \\ & +0.042(\pm 0.006)Z4_{H_2O,H^+} + 0.099(\pm 0.011)Z5_{H_2O,H^+} + 0.031(\pm 0.006)Z6_{H_2O,H^+} \\ & + 0.054(\pm 0.012)Z7_{H_2O,H^+} + 7.665(\pm 0.026) \\ & n = 48,\ s = 0.182,\ r = 0.977,\ F = 123.0,\ P = 0.0001,\ \text{press } s = 0.409 \end{aligned} \quad (5)$$

The standard error of estimation reduced from 0.383 in Eq. 2 to 0.182 in Eq. 5, and the correlation coefficient increased from 0.885 to 0.977. The press s value from the leave-one-out cross-validation test is reasonable. The observed and calculated biological activity using Eq. 5 and the latent variables used are given in Table 4. Table 5 is the summary of press and standard error of estimation values from the cross-validation tests. The results of the cross-validation tests show that further reduction of the standard error is possible by additional latent variables. In this study, however, the seven-latent-variable model was conservatively chosen considering both the reduction of PRESS values and the stepwise F-statistics of the model. A larger number of excluded compounds in the jackknife cross-validation tests did not significantly affect the results as summarized in Table 6.

An inclusion of the electrostatic effects with the steric effects (CH_3 probe) did not improve the correlation. In fact, the standard error or estimation of the fitted three-latent-variable model increased from 0.426 of the CH_3 probe to 0.442, and the correlation coefficient reduced from 0.856 to 0.844. The standard error of estimations of the cross-validation tests for CH_3 alone or CH_3 and H^+ probes together for the three-variable model remained essentially unchanged (Table 5). The overall quality of the correlation from H_2O and H^+ probes was considerably better than that from CH_3 and H^+.

Figures 2 and 3 are the coefficient contour plots of the correlation described in Eq. 5. The hydrophobic contours (Fig. 2) were drawn at the 0.02 level and the electrostatic contours (Fig. 3) at the 0.015 level. These contours correspond to the lattice points whose products between the 3D-QSAR coefficients of Eq. 5 and the associated standard deviation values of the original

TABLE 4
OBSERVED AND CALCULATED Log $1/IC_{50}$ VALUES USING EQ. 5

No.	Compound	Log $1/IC_{50}$			PLS latent variables						
		obs	calc	dev	$Z1_{H_2O.H^+}$	$Z2_{H_2O.H^+}$	$Z3_{H_2O.H^+}$	$Z4_{H_2O.H^+}$	$Z5_{H_2O.H^+}$	$Z6_{H_2O.H^+}$	$Z7_{H_2O.H^+}$
1	RO-054318	6.34	6.14	0.20	−10.43	−4.044	−3.261	−3.279	−0.853	5.763	−3.029
2	RO-053072	6.41	6.40	0.01	−10.78	−2.494	−1.829	−6.159	2.041	1.767	−1.726
3	RO-054528	6.42	6.93	−0.51	0.927	−4.636	0.834	0.989	−3.675	−0.723	−2.566
4	RO-052921	6.45	6.85	−0.40	−9.035	−1.513	−0.318	−3.987	1.053	0.457	2.150
5	RO-207736	7.02	7.02	−0.00	−2.362	0.296	−0.112	−3.289	−1.680	3.830	−5.134
6	RO-054619	7.12	7.25	−0.13	−6.185	2.611	0.890	−6.479	2.345	1.334	−2.564
7	RO-205397	7.37	7.66	−0.29	0.054	−2.342	1.578	−2.926	1.510	0.406	1.422
8	RO-053061	7.40	7.28	0.12	−5.380	−3.102	0.044	−2.942	2.210	1.398	3.144
9	RO-202533	7.44	7.50	−0.06	−6.232	2.284	4.241	−0.034	0.447	−1.765	0.284
10	RO-202541	7.52	7.73	−0.21	3.971	0.100	2.863	1.091	−3.437	−0.081	−1.771
11	RO-205747	7.62	7.78	−0.16	−5.459	2.681	5.575	1.656	1.409	−1.558	−1.376
12	RO-054336	7.68	7.55	0.13	−6.333	2.787	1.682	−3.936	1.421	0.815	2.351
13	RO-203053	7.74	7.75	−0.01	0.976	2.956	5.810	0.199	0.096	−5.928	−5.359
14	RO-052904	7.89	8.22	−0.33	1.687	−1.700	3.868	1.482	1.825	−1.019	3.344
15	Diazepam	8.09	8.04	0.05	−2.598	−1.559	2.569	4.831	0.138	8.000	2.869
16	RO-116896	8.15	8.09	0.06	11.954	−4.364	−2.892	−1.394	−0.379	2.346	−1.737
17	RO-075220	8.26	8.30	−0.04	2.373	2.951	3.510	3.169	−2.534	3.779	1.132
18	RO-143074	8.27	8.31	−0.04	−3.476	4.499	5.459	−0.636	5.120	2.293	−3.350
19	Flunitrazepam	8.42	8.26	0.16	9.367	−4.352	1.359	0.792	−0.397	3.078	0.231
20	RO-053590	8.45	8.39	0.06	11.515	0.486	3.123	−5.569	−0.864	−4.671	1.177
21	RO-079957	8.54	8.30	0.24	−0.818	5.765	6.150	4.515	−2.042	1.890	−0.701
22	RO-133780	8.62	8.62	−0.00	0.097	3.781	5.653	5.788	−0.942	4.999	2.257
23	RO-053367	8.70	8.52	0.18	0.375	4.509	5.968	0.594	1.463	−0.843	1.691
24	Delorazepam	8.74	8.57	0.17	2.245	4.956	6.523	−0.294	1.082	−1.838	0.503

TABLE 4 (continued)

No.	Compound	Log $1/IC_{50}$			PLS latent variables						
		obs	calc	dev	$Z1_{H_2O,H^+}$	$Z2_{H_2O,H^+}$	$Z3_{H_2O,H^+}$	$Z4_{H_2O,H^+}$	$Z5_{H_2O,H^+}$	$Z6_{H_2O,H^+}$	$Z7_{H_2O,H^+}$
25	Clonazepam	8.74	8.72	0.02	10.315	−2.847	3.271	−2.584	3.243	0.124	1.261
26	RO-054435	8.82	8.56	0.26	8.663	−3.366	2.654	−2.512	3.068	0.560	2.048
27	Meclonazepam	8.92	8.90	0.02	13.661	−2.204	0.278	−3.552	3.952	0.606	−0.077
28	RO-053328	7.06	7.12	−0.06	−1.762	2.160	1.693	−4.363	−5.404	−5.586	4.585
29	Tetrazepam	7.47	7.52	−0.05	−3.128	−0.776	1.267	1.458	−4.139	5.520	5.635
30	Bromazepam	7.74	7.50	0.24	−4.208	−0.229	3.826	0.442	0.732	−2.167	0.579
31	Demoxepam	6.51	6.40	0.11	−7.038	−5.185	−0.408	−0.210	−0.005	−5.757	−1.291
32	Chlordiazepoxide	6.45	6.41	0.04	−4.799	−12.04	−5.486	9.358	2.336	−9.279	−0.348
33	Alprazolam	7.70	7.59	0.11	−2.027	0.192	−1.676	5.435	−0.963	4.991	−2.503
34	Triazolam	8.40	8.57	−0.17	2.602	5.629	1.285	5.354	−0.413	5.290	−2.812
35	RO-116679	8.40	8.68	−0.28	11.057	−3.093	−5.553	1.860	2.266	9.575	−0.832
36	RO-221892	7.92	8.03	−0.11	−0.723	9.563	−16.43	−2.794	3.125	−0.094	2.644
37	RO-215205	8.13	8.07	0.06	2.048	13.034	−12.90	−3.470	−0.786	−2.251	0.315
38	Midazolam	8.32	8.35	−0.03	2.049	6.316	−6.470	6.034	0.182	0.874	−0.375
39	RO-141359	7.15	7.20	−0.05	2.326	2.449	1.372	−4.914	−4.245	−7.859	−0.753
40	RO-158867	7.60	7.61	−0.01	12.821	−4.223	−0.698	−6.974	−2.357	−6.844	−0.654
41	RO-147187	6.39	6.15	0.24	−6.319	−2.464	−7.916	−3.344	−4.448	4.488	−0.613
42	RO-159270	8.30	8.10	0.20	12.434	−3.557	−3.929	−1.548	−2.026	4.960	−1.132
43	Desmethylclobazam	6.68	6.68	0.00	−6.067	−4.136	−0.427	0.200	0.139	−4.371	−0.427
44	Clobazam	6.89	6.84	0.05	−5.877	−5.951	−1.631	4.003	−1.301	4.127	0.535
45	RO-221251	7.96	7.95	0.01	2.673	0.635	−4.881	10.687	−0.077	−8.340	0.404
46	RO-223245	8.55	8.48	0.07	4.579	4.571	−2.880	8.499	1.191	−7.650	−0.868
47	RO-223148	6.38	6.36	0.02	−7.556	−7.248	−5.899	1.075	0.782	−2.645	1.636
48	Premazepam	6.77	6.67	0.10	−12.18	2.212	2.252	−2.320	−0.208	−1.998	−1.201

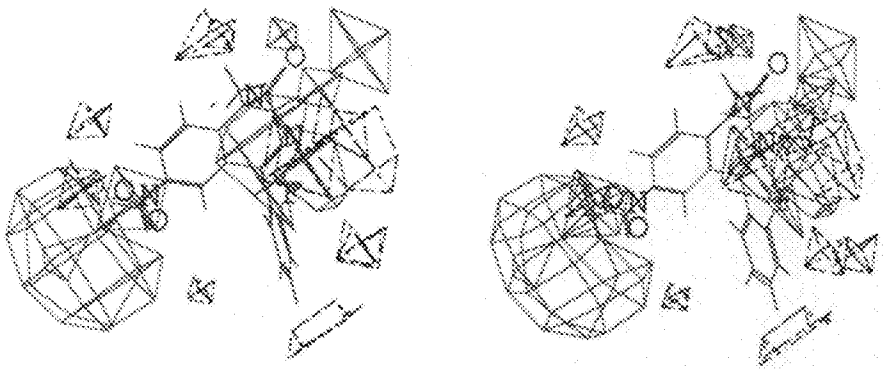


Fig. 2. Coefficient contour plot of the correlation described in Eq. 5. The areas where hydrophobic substituents increase the binding affinity are shown by the positive contours in red (solid line) and the areas where they decrease it are shown by the negative contours in blue (dashed line). The contours were drawn at the 0.02 level. Flunitrazepam (**19**) was used as a reference compound.

variables were interpolated as +0.02 or +0.015 (positive contours) and −0.02 or −0.015 (negative contours), respectively. Flunitrazepam (**19**) was shown as a reference compound. The areas where hydrophobic substituents increase the binding affinity are shown by the positive contours in red (solid lined) and the areas where they decrease it are shown by the negative contours in blue (dashed lined). The positive hydrophobic contours around the 7-position are consistent with the positive coefficient of the $\pi 7$ variable in the traditional QSAR described in Eq. 1. The results are in contrast to the previous study with SYBYL-CoMFA where the hydrophobic effects of 7-substituents could not be represented in the coefficient contour plot. Additional positive hydrophobic contours are located around the 2′-position. However, it is not possible to state conclusively at the present time that the positive contours around 2′-position represent the hydrophobic effects, since the structural variance at this position is rather limited. Nonetheless, the result is consistent with the positive coefficient of the $I_{2'}$ parameter in Eq. 1, and indicates that hydropho-

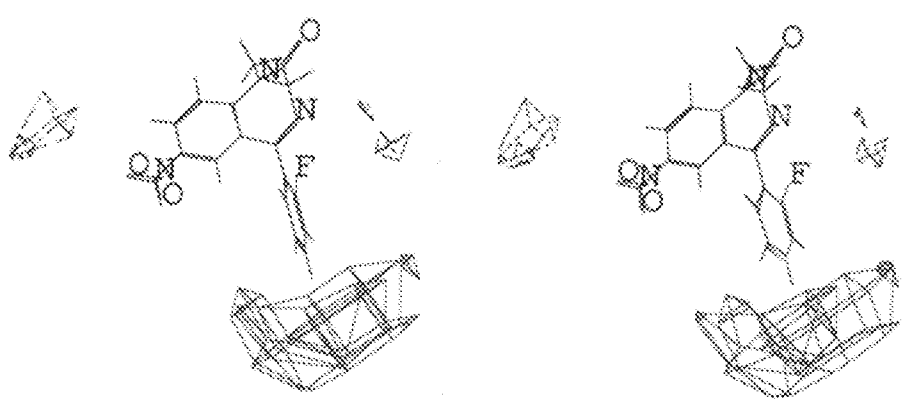


Fig. 3. Coefficient contour plot of the correlation described in Eq. 5. The areas where the electrostatic effects influence the binding affinity are shown by the positive contours in purple (solid line) and the negative contours in cyan (dashed line). The contours were drawn at the 0.015 level. Flunitrazepam (**19**) was used as a reference compound.

TABLE 5
SUMMARY OF PRESS FROM LEAVE-ONE-OUT CROSS-VALIDATION TESTS

	L	PRESS	press s	Average error
	0	31.1360	0.8054	0.688
(a) H_2O probe	1	17.0115	0.5953	0.497
	2	14.3469	0.5467	0.459
	3	12.1165	0.5024	0.400
	4	11.5739	0.4910	0.363
	5	11.4987	0.4894	0.379
	6	11.5114	0.4897	0.389
	7	10.2087	0.4612	0.363
	8	9.7317	0.4503	0.352
	9	8.8459	0.4293	0.331
	10	8.0734	0.4101	0.325
	11	7.8654	0.4048	0.327
	12	7.7563	0.4020	0.330
	13	8.6668	0.4249	0.350
(b) H_2O plus H^+ probe	1	18.9517	0.6284	0.525
	2	13.3841	0.5280	0.444
	3	11.7096	0.4939	0.403
	4	11.1014	0.4809	0.363
	5	11.6147	0.4919	0.406
	6	10.0454	0.4575	0.359
	7	8.0413	0.4093	0.328
	8	6.7353	0.3746	0.303
	9	5.5662	0.3405	0.281
	10	5.1235	0.3267	0.271
	11	4.6350	0.3107	0.261
	12	4.3848	0.3022	0.254
	13	4.4978	0.3061	0.255
(c) CH_3 probe	1	17.0605	0.5962	0.486
	2	15.6904	0.5717	0.476
	3	14.2245	0.5444	0.436
	4	15.4576	0.5675	0.473
(d) CH_3 and H^+ probe	1	19.4118	0.6359	0.526
	2	15.4506	0.5674	0.475
	3	14.0851	0.5417	0.445
	4	14.9438	0.5580	0.461
(e) H^+ probe	1	23.5705	0.7008	0.584
	2	20.1615	0.6481	0.532
	3	19.3544	0.6350	0.548
	4	16.0677	0.5786	0.480
	5	16.2833	0.5824	0.483
(f) H^+ probe (no steric cutoff)	1	21.7678	0.6734	0.562
	2	19.1038	0.6309	0.526
	3	19.2810	0.6338	0.540
	4	18.3645	0.6185	0.533
	5	17.0794	0.5965	0.507
	6	17.0716	0.5964	0.502
	7	16.6933	0.5897	0.490
	8	17.8683	0.6101	0.516

TABLE 6
SUMMARY OF STATISTICS FROM LEAVE-N-OUT CROSS-VALIDATION TESTS FOR EQ. 5 (L = 7)

N	PRESS	Press s	Average error
1	8.0413	0.4093	0.328
2	8.3849	0.4180	0.331
3	8.9712	0.4323	0.344
4	8.2463	0.4145	0.332
5	7.8671	0.4048	0.330
6	10.5209	0.4682	0.371
7	7.4478	0.3939	0.333
8	7.4478	0.3939	0.333
9	10.3479	0.4643	0.391
10	10.2290	0.4616	0.363
15	8.8423	0.4292	0.348
20	10.8514	0.4755	0.389

bic groups are preferred for improving the affinity at this position. $I_{2'}$ in Eq. 1 is an indicator variable and holds the value of 1.0 if the 2′-position carries a substituent such as Cl, F, and CF_3 and 0.0 when the 2′-position is not substituted. The areas where the electrostatic effects influence the binding affinity are shown by the positive in purple (solid lines) and negative contours in cyan (dashed lines). The positive electrostatic contours around the 4′-position of the 5-phenyl ring indicate that decreased electron density on the phenyl ring, which is related to an electron-withdrawing effect of the 2′-substituent, would increase the binding affinity. Small negative electrostatic contours are located near the 7-position. An electron-withdrawing group at this position increases the binding affinity.

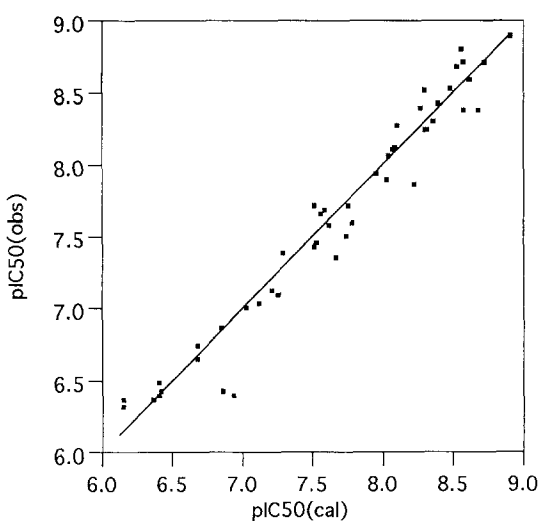


Fig. 4. Observed vs. calculated pIC_{50} values using Eq. 5.

TABLE 7
RELATIVE CONTRIBUTION OF THE MOLECULAR DESCRIPTORS TO THE CORRELATION

Effects	QSAR	SYBYL-CoMFA ($100 \times \pi_7$ and H^+ probe)	GRID-CoMFA (H_2O and H^+ probes)[a]
Electrostatic	25% (E_{HOMO})	70% (H^+)	18% (H^+)
Hydrophobic	29% (π_7)	30% (π_7)	78% (H_2O)
Others	46% ($I_{2'}$)	–	–

[a] Normalized percentage is 19% (H^+) and 81% (H_2O), respectively.

Figure 4 is the plot of the observed vs. calculated pIC_{50} values using Eq. 5. Table 4 lists the observed and calculated pIC_{50} values of the compounds using this equation.

The current GRID-CoMFA results are in general comparable to the traditional QSAR (Eq. 1) that Greco et al. [12] described previously. The QSAR of both the homogeneous subset (first 30 out of 48 compounds) and the heterogeneous set (all 48 compounds) indicate that π_7 explains more variance in the biological data than the electronic parameters (E_{LUMO} or E_{HOMO}). The results of this GRID-CoMFA study suggest that the hydrophobic effects alone explain the majority (78%) of the variance in the data. In the GRID-CoMFA, the H_2O probe also explains the portion accounted for by the I_2 indicator variable in QSAR. The electrostatic effects explain an additional 18% of the variance. However, since the hydrogen bonding effects accounted for by the H_2O probe partially include the electrostatic effects described with the H^+ probe, it is difficult to accurately evaluate the actual contributions of hydrophobicity and electrostatics in modulating the affinity. The SYBYL-CoMFA results show 70% of the electrostatic effects and 30% of the hydrophobic effects. It appears that the portion explained by the indicator variable $I_{2'}$ (46%) in Eq. 3 is accounted for by the electrostatic effects in the SYBYL-CoMFA model (See Table 7).

The current CoMFA model obtained using the GRID force field improved the correlation significantly over those of the traditional QSAR or the CoMFA obtained using the SYBYL force field. The standard deviation and correlation coefficient of the fitted model from GRID-CoMFA are 0.182 and 0.977, while those of QSAR are 0.410 (0.392) [27] and 0.867, respectively (Eq. 1). The corresponding statistical indices obtained by the SYBYL-CoMFA were quite close (s = 0.401 (0.388) [27] and r = 0.871) to those of QSAR. Those from the H_2O alone in the GRID-CoMFA are 0.383 and 0.885, respectively.

In the SYBYL-CoMFA the combined use of the steric and electrostatic fields gave a worse

TABLE 8
SQUARED CORRELATION COEFFICIENTS MATRIX OF THE INDEPENDENT VARIABLES USED IN THE QSAR ANALYSIS

	π_7	$I_{2'}$	E_{HOMO}	E_{LUMO}
π_7	1.00	0.03	0.00	0.03
$I_{2'}$		1.00	0.11	0.24
E_{HOMO}			1.00	0.64
E_{LUMO}				1.00

TABLE 9
STATISTICS OF THE LATTICE POINTS FOR H_2O, H^+, AND CH_3 PROBES USED IN GRID-CoMFA

Probe	No. of points	No. of total points[a]	Mean	Standard deviation	Minimum value	Maximum value
H_2O	221	10 608	−0.176	1.470	−4.280	4.000
H^+	239	64 757	−0.014	0.353	−3.638	2.337
CH_3	1349	11 472	0.042	1.365	−2.716	4.000

[a] Number of points × number of compounds (48).

result than the steric field alone (one component model, press s = 0.599). Similar results were obtained with the GRID-CoMFA where the three-component CH_3 model gives a best result (press s = 0.544) and the addition of electrostatic field did not change the results (three-component model, press s = 0.542). Although the improvement is significant, a higher number of components was required in the GRID-CoMFA model. It is evident that the SYBYL C_{SP_3} probe with a +1.0 charge is not completely equivalent to the CH_3 plus H^+ probes in the GRID-CoMFA model.

In CoMFA, before the computed interaction energy values were subjected to the data analysis by PLS, they were reduced to a smaller number by eliminating points that vary little or that have high-energy values. In the process of reducing the data matrix in this study, one major modification was made from the original SYBYL-CoMFA methodology. This involved the lattice points that are outside the molecule but inside some other molecules included in the study. In the original CoMFA procedures, the mean value of the other molecules' electrostatic interactions at the same location is used for such lattice points. In the GRID-CoMFA procedures, these points were totally excluded from consideration. When the molecules are placed into the receptor cavity, it appears to us that these points are not likely to represent the charged group of a receptor. The results that included these points (no steric cutoff) were less good as is shown in Table 5 (f).

In the validation tests, a different number of compounds can be excluded for prediction. The number of cross-validation groups in previous work [12] was 10 (leave-five-out), while it was 48 (leave-one-out) in the present GRID-CoMFA analysis. Thus the two results are not directly comparable. However, the results of cross-validation, in terms of PRESS values, were found relatively insensitive to the variation of the number of excluded and predicted observations in the cross-validation test (see Table 6).

There were no significant correlations among the independent variables used in Eq. 1 of the traditional QSAR (See Table 8). On the other hand, the CoMFA models from the GRID force field or the SYBYL force field suffer from some collinearity problems with the current set of compounds for their interpretations. In the SYBYL-CoMFA model the electrostatic descriptor does not discriminate between the electron effects of the 7-substituent and the unidentified effects of the 2′-substituent. In the GRID-CoMFA model, it is not clear at this point how important the electrostatic effects of the 7-substituent are with respect to its hydrophobic effects. Table 9 summarizes the statistics of the interaction energies at the total lattice points for H_2O, CH_3, and H^+ probes used in this study.

CONCLUSION

The results presented in this study demonstrate that the use of the H_2O probe with the hydrogen bond potential in the GRID-CoMFA has the advantage of describing hydrophobic effects without any bias unlike the QSAR-CoMFA approach reported previously. As shown in this study, the correlation suggesting the importance of the hydrophobic and electrostatic effects was obtained without prior knowledge or hypotheses of their effects. Thus, it would have been possible to obtain information about the presence of hydrophobic effects even if no traditional QSAR had previously been derived.

The usefulness of including hydrophobicity in CoMFA as described in this study is evident in the literature when one considers that hydrophobic effects often play one of the critical roles influencing the binding between ligands and receptors. Furthermore, the use of log P in the traditional QSAR studies is often limited by the availability of proper values for the investigated compounds.

REFERENCES AND NOTES

1 Martin, Y.C., In Quantitative Drug Design: A Critical Introduction (Medicinal Research Series, Vol. 8), Marcel Dekker, New York, Basel, 1978.
2 Cramer, III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
3 Ståhle, L. and Wold, S., Prog. Med. Chem., 25 (1988) 292.
4 Lindberg, W., Persson, J.-A. and Wold, S., Anal. Chem., 55 (1983) 643.
5 Hoskuldsson, A., J. Chemometrics, 2 (1987) 211.
6 Cramer, III, R.D., Bunce, J.D., Patterson, D.E. and Frank, I.E., Quant. Struct.-Act. Relat., 7 (1988) 18.
7 Wold, S., Technometrics, 20 (1978) 397.
8 Allen, M.S., Tan, Y., Trudell, M.I., Narayanan, K., Schindler, I.R., Martin, M.J., Schultz, C., Hagen, T.J., Koehler, K.F., Codding, P.W., Skonick, P. and Cook, J.M., J. Med. Chem., 33 (1990) 2343.
9 Bjorkroth, J.-P., Pakkanen, T.A., Lindroos, J., Pohjala, E., Hanjijarvi, H., Lauren, L., Hannuniemi, R., Juhakoshi, A., Kippo, K. and Kleimola, T., J. Med. Chem., 34 (1991) 2338.
10 Carroll, F.I., Gao, Y., Rahman, A., Abraham, P., Parham, K., Lewin, A.H., Boja, J.W. and Kuhar, M.J., J. Med. Chem., 34 (1991) 2719.
11 Greco, G., Novellino, E., Silipo, C. and Vittoria, A., Quant. Struct.-Act. Relatsh., 10 (1991) 289.
12 Greco, G., Novellino, E., Silipo, C. and Vittoria, A., Quant. Struct.-Act. Relatsh., 11 (1992) 461.
13 Clark, M., Cramer, III, R.D., Lones, D.M., Patterson, D.E. and Simeroth, P.E., Tetrahedron Comp. Method., 3 (1990) 47.
14 Kim, K.H. and Martin, Y.C., J. Org. Chem., 56 (1991) 2723.
15 Kim, K.H. and Martin, Y.C., J. Med. Chem., 34 (1991) 2056.
16 Kim, K.H., In Wermuth, C.G. (Ed.) Trends in QSAR and Molecular Modelling '92, ESCOM Leiden, 1993, in press.
17 Kim, K.H., Med. Chem. Res., 1 (1991) 259.
18 Goodford, P.J., J. Med. Chem., 28 (1985) 849.
19 Boobbyer, D.N.A., Goodford, P.J., McWhinnie, P.M. and Wade, R.C., J. Med. Chem., 32 (1989) 1083.
20 GRID Program, V7, Molecular Discovery Ltd., West Way House, Elms Parade, Oxford, U.K.
21 SYBYL Molecular Modelling System (Version 5.30), TRIPOS Assoc., St. Louis, MO, U.S.A.
22 Hansch, C. and Leo, A., In Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, 1979.
23 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., J. Am. Chem. Soc., 107 (1985) 3902.
24 MOPAC (Version 5.00), Quantum Chemistry Program Exchange, No. 455, 1989.
25 The terms GRID-CoMFA and SYBYL-CoMFA are used throughout this paper in order to distinguish the different

force fields (GRIDS and SYBYL force fields, respectively) used in the CoMFA study. The term QSAR-CoMFA refers to the study where both the traditional QSAR and CoMFA methodologies are employed in order to investigate structure–activity relationships of a set of compounds.

26 Haefely, W., Kyburz, E., Gerecke, M. and Moheler, H., In Testa, B. (Ed.) Advances in Drug Researches, Vol. 14, Academic Press, London, 1985, pp. 156–322.

27 The originally reported standard error of the calibration was transformed into the root mean square of estimation calculated as described in the Methods section and reported in parentheses in order to compare the results.