3D-Pharmacophores of Flavonoid Binding at the Benzodiazepine $GABA_A$ Receptor Site Using 4D-QSAR Analysis

Xuan Hong and A. J. Hopfinger*

Laboratory of Molecular Modeling and Design (M/C 781), College of Pharmacy, The University of Illinois at Chicago, 833 South Wood Street, Chicago, Illinois 60612-7231

Received May 20, 2002

4D-QSAR analysis was applied to a training set of 38 flavonoids where affinity constants, Ki, to the $GABA_A$ benzodiazepine receptor site, BzR, were determined. It was found that the −logKi values of the compounds are highly dependent on the size and electrostatics character of the substituents at the $R_{3'}$ and R_6 positions of the flavonoid scaffold. Polar negative groups correctly embedded in the $R_{3'}$ and/or R_6 substituents are predicted to increase −logKi values. A planar conformation of the flavonoid scaffold was found not to be a requirement for the flavonoids to be active. A test set of four compounds was used to evaluate the predictivity of the 4D-QSAR models.

INTRODUCTION

GABA (γ-aminobutyric acid) is a major inhibitory transmitter in the central nervous system (CNS). GABA is also involved as a neurotransmitter and/or a paracrine effector in the regulation of a variety of physiological mechanisms in the periphery.[1] The GABA receptor system has three major types: $GABA_A$, $GABA_B$ and $GABA_C$. $GABA_A$ receptors, which are the ligand-gated chloride ion channel complexes, are important in the mammalian synapses of the CNS. The BZDs (benzodiazepines), a family of psychoactive drugs that are anxiolytic, anticonvulsant, muscle relaxant and sedative-hypotic, exert their actions by binding to $GABA_A$ receptors, causing an "allosteric modification" of the GABA binding site, and making GABA binding more effective.[2] The BZDs are considered the most safe psychotropic drugs available today. However, they still have a series of unwanted side effects, such as sedative and myorelaxant actions, ethanol and barbiturate potentiation, recorded amnesia, ataxia and the potential for drug abuse and tolerance.[3]

Studies[4−7] have shown that some naturally occurring flavonoids, and synthetic flavone derivatives, also have high binding affinity to the BZD receptor (BzR), while devoid of unwanted side effects. Therefore, the flavonoid family may provide important leads for the development of potent and selective BzR ligands that could also be free of possible disadvantages. Efforts have gone into finding a pharmacophore model for the binding of flavonoids to the BzR of the $GABA_A$ receptors.[8−11] To study flavonoid-BzR interactions more thoroughly, 4D-QSAR[12] analysis has been applied to a data set of flavonoid ligands to BzR. 4D-QSAR analysis has been successfully applied to a variety of structure−activity training sets[13−19] and has proved useful and reliable in the construction of quantitative 3D pharmacophore models when the geometry of the corresponding receptor is not known. The merits of the 4D-QSAR methodology are in its ability to a) incorporate ligand conformational flexibility, b) multiple alignment exploration, and c) exhaustive evaluation of ligand-embedded pharmacophore groupings as part of the QSAR model building and optimization process.

METHODS

1. The Training Set of Flavonoids. A series of 38 flavonoids, chosen from the literature,[10] was used as the training set. The chemical structures of the flavonoids and their BzR site binding affinities, represented by −logKi, are given in Table 1. Ki is the ligand−receptor affinity constant. The last four compounds given in Table 1 were not included in 4D-QSAR model development but rather employed as a test set.

2. Receptor-Independent 4D-QSAR Analysis Applied to the Flavonoid Training Set. The 4D-QSAR scheme can be applied to either receptor-dependent (RD) or receptor-independent (RI) training sets. In the first scheme, a form of structure-based design, the geometry of the receptor (or molecular target) is available. In contrast, in the second scheme, either the geometry of the receptor is not available, or it is not used in the 4D-QSAR analysis because of uncertainty in the receptor geometry and/or the binding alignment of the ligand to the receptor. The RI 4D-QSAR formalism was applied in this study since the geometry of the BzR on the $GABA_A$ receptor is unknown.

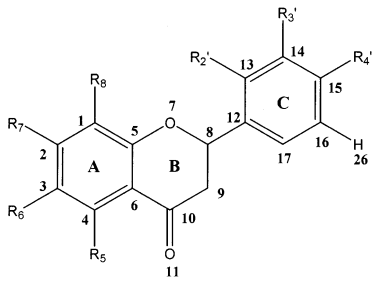
The current formulation of RI 4D-QSAR analysis consists of 10 operational steps that have been described in detail previously[12−16] and are summarized here only in terms of specific modeling of the flavonoid training set.

Step 1. A three-dimensional structure of each of the 38 flavonoids given in Table 1 was constructed in the neutral form using the *HyperChem 5.01* software.[20] Partial atomic charges were computed using the AM1 semiempirical method also as implemented in the *HyperChem* program. Each structure was energy-minimized using the *HyperChem 5.01* quantum mechanical method without any geometric constraint. The energy minimized structures were used as the initial structures in conformational sampling, see **Step 3** below.

* Corresponding author phone: (312)996-4816; fax: (312)413-3479; e-mail: hopfingr@uic.edu.

10.1021/ci0200321 CCC: $25.00
Published on Web 12/05/2002

Table 1. Chemical Structures of the Flavonoids Used in the 4D-QSAR Analyses and the Corresponding BzR Site Binding Affinities (−logKi, Ki in uM) and ClogP Values of Each Flavonoid



compound	R_5	R_6	R_7	R_8	$R_{2'}$	$R_{3'}$	$R_{4'}$	−logKi	ClogP[a]
1	−H	−H	−H	−H	−H	−H	−H	6.00	2.32
2	−H	−F	−H	−H	−H	−OH	−H	5.60	2.17
3	−H	−Cl	−H	−H	−H	−OH	−H	6.07	2.55
4	−H	−Br	−H	−H	−H	−OH	−H	6.22	2.83
5	−H	−F	−H	−H	−H	$-NO_2$	−H	6.74	2.41
6	−H	−Cl	−H	−H	−H	$-NO_2$	−H	8.10	2.79
7	−H	−Cl	−H	−H	−H	−H	$-OCH_3$	5.90	2.58
8	−H	−Br	−H	−H	−H	−H	$-OCH_3$	5.68	2.86
9	−H	−Br	−H	−H	$-NO_2$	−H	−H	6.68	3.06
10	−H	$-NO_2$	−H	−H	−H	−H	−Br	7.60	3.06
11	−H	−Cl	−H	−H	−F	−H	−H	6.38	2.98
12	−H	−Br	−H	−H	−F	−H	−H	6.42	3.25
13	−H	−H	−H	−H	−H	−F	−H	5.45	2.46
14	−H	−F	−H	−H	−H	−F	−H	6.04	2.60
15	−H	−Cl	−H	−H	−H	−F	−H	6.93	2.98
16	−H	−Br	−H	−H	−H	−F	−H	7.38	3.25
17	−H	−H	−H	−H	−H	−H	−F	5.44	2.46
18	−H	−F	−H	−H	−H	−H	−F	5.60	2.60
19	−H	−Cl	−H	−H	−H	−H	−F	6.74	2.98
20	−H	−Br	−H	−H	−H	−H	−F	6.94	3.25
21	−H	−H	−H	−H	−H	−Cl	−H	6.21	2.84
22	−H	−F	−H	−H	−H	−Cl	−H	6.70	2.98
23	−H	−Cl	−H	−H	−H	−Cl	−H	7.64	3.35
24	−H	−Br	−H	−H	−H	−Cl	−H	7.77	3.63
25	−H	−H	−H	−H	−H	−Br	−H	6.38	3.25
26	−H	−F	−H	−H	−H	−Br	−H	6.63	3.25
27	−H	−Cl	−H	−H	−H	−Br	−H	7.64	3.63
28	−H	−Br	−H	−H	−H	−Br	−H	7.72	3.90
29	−H	−Br	−H	−H	−H	−H	−H	7.15	3.11
30	−H	−Br	−H	−H	−H	−H	$-NO_2$	6.70	3.06
31	−H	$-NO_2$	−H	−H	−H	$-NO_2$	−H	7.92	2.23
32	−H	−Br	−H	−H	−H	$-NO_2$	−H	9.00	3.06
33	−OH	−Br	−OH	−Br	−H	−H	−H	6.15	3.33
34	−OH	−H	−OH	−H	−H	−H	−H	5.52	1.75
35	−OH	−H	−OH	−H	−H	−H	−OH	5.52	1.46
36	−OH	−H	−OH	−H	−Cl	−H	−H	5.10	2.27
37	−OH	−H	−OH	−H	−F	−H	−H	5.10	1.89
38	−OH	$-OCH_3$	−OH	−H	−H	−H	−OH	6.00	1.21
baicalin[b]	−OH	−OH	$-C_6H_9O_7$[c]	−H	−H	−H	−H	4.11	0.11
baicalein[b]	−OH	−OH	−OH	−H	−H	−H	−H	5.25	1.46
scutellarein[b]	−OH	−OH	−OH	−H	−H	−H	−OH	4.92	1.18
wogonin[b]	−OH	−H	−OH	$-OCH_3$	−H	−H	−H	5.69	1.50

[a] ClogP was calculated from the ChemPlus program of the HyperChem software suite. [b] Test compounds. [c] $-C_6H_9O_7$ = 5-carboxyglucosyloxy.

Step 2. The atoms of each flavonoid were classified into seven types of *interaction pharmacophore elements, IPEs*, which are defined in Table 2. This IPE classification scheme is used to define the type of interaction for each pharmacophore site of a 4D-QSAR model.

Step 3. Molecular dynamic simulation, MDS, was used to sample the conformational states available to each flavonoid and to generate its corresponding *conformational ensemble profile* (CEP). The MDSs were performed using the *MOLSIM* package[21] with an extended MM2 force field.[22,23] The temperature for the MDS was set at 300 K with a simulation sampling time of 40 picoseconds (ps) with intervals of 0.001 ps for a total sampling of 40 000 conformations of each flavonoid. The atomic coordinates of

Table 2. Interaction Pharmacophore Elements, IPEs, Used in the 4D-QSAR Study

IPE description	IPE symbol
any type of atom	*any (0)*
nonpolar atom	*np (1)*
polar atom of partial positive charge	*p+ (2)*
polar atom of partial negative charge	*p- (3)*
hydrogen bond acceptor	*hba (4)*
hydrogen bond donor	*hbd (5)*
aromatic carbon and hydrogen	*ar (6)*

Table 3. Sets of Trial Alignments Used in Constructing of the 4D-QSAR Models[a]

alignment sets		1st atom	2nd atom	3rd atom
set 1 (whole)	1	2	11	15
	2	3	7	15
	3	1	11	14
	4	4	11	17
	5	R_6	$R_{3'}$	26
set 2 (right)	6	9	8	12
	7	14	15	16
	8	8	12	13
set 3 (left)	9	1	2	3
set 4 (left & middle)	10	2	7	11
set 5 (right & middle)	11	7	11	15

[a] See Table 1 for reference atom coding. "Whole" means the alignments span the "whole" structure (i.e. ring A, B, and C) of the flavonoids. "right" means the alignments only cover the "right" part of the structures (i.e. ring C), while the alignments under the "left", "left & middle" and "right & middle" categories cover ring A, A&B and B&C, respectively.

each conformation and its corresponding intramolecular energy sampled during the MDS were recorded every 0.02 ps for a total of 2000 "frames", or steps, in the CEP of each flavonoid.

Step 4. The current (RI) 4D-QSAR methodology uses 3-ordered atom alignments to compare the molecules of a training set. Eleven alignments that span the entire flavone structure were selected in this study and are defined in Table 3.

Step 5. Each conformation of an analogue from its CEP is aligned relative to a grid-cell lattice through the invariant coordinates of the three-ordered alignment atoms. In this study the size of a grid cell is 1 $Å^3$, and the dimension of the overall grid cell lattice was chosen to completely embed each flavonoid of the training set. The normalized occupancy of each grid cell by each IPE atom type over the CEP of each molecule, for a given alignment, forms a unique set of QSAR descriptors referred to as *grid cell occupancy descriptors*, GCODs. The GCODs are computed and used as the basis set of trial 4D-QSAR descriptors in a 4D-QSAR analysis.

Step 6. A 4D-QSAR analysis generates an enormous number of trial QSAR descriptors, GCODs, because of the large number of grid cells and the seven IPEs. Two serial levels of data reduction can be considered. The first level of data reduction eliminates GCODs for which their variance (self-variance) over a training set is less than a prechosen fraction or percentage. In this study, a less than 2% variation was used as a reduction filter. All the remaining GCODs are mean-centerized. Partial least-squares (PLS) regression analysis[24] is then performed as a data reduction fit between the observed dependent variable (in this study, the binding affinities of analogues to the BzR in the $GABA_A$ receptors) measures and the corresponding set of GCOD values.

Step 7. The M (in this study 200) most highly weighted PLS GCOD descriptors, generated in Step 6, are used to form the trial basis set for a genetic algorithm (GA) model optimization. The specific genetic algorithm currently used in the 4D-QSAR software[25] is a modification of the genetic function approximation, GFA.[26,27] The GFA optimization is initiated using N (currently 300) randomly generated 4D-QSAR models. Mutation probability over the crossover optimization cycle is set at 10%. The smoothing factor, a GFA control variable, which specifies the number of independent variables in the QSAR models, is varied in order to determine the optimal number of descriptors in the 4D-QSAR models based on multidimensional linear regression, MLR, data fitting. The diagnostic measures used to analyze the resultant 4D-QSAR models generated by the GFA include the following: 1. descriptor usage as a function of crossover operation, 2. linear cross-correlation among descriptors and/or dependent variables (biological activity measures), 3. number of significant and independent 4D-QSAR models, and 4. indices of model significance including the correlation coefficient, r^2, "leave-one-out" cross-validation correlation coefficient, (xv-r^2), and Friedman's lack of fit, LOF.[28]

Non-4D-QSAR descriptors can also be included in the GFA model optimization. In this study, the logarithm of the 1-octanol/water partition coefficient, logP, was added in the form of ClogP which was calculated using the *ChemPlus* program implemented as part of the *HyperChem* software. The "C" in ClogP indicates this property is calculated as opposed to measured. The ClogP values for each of the 38 flavonoids are given as part of Table 1. The purpose of adding logP as a trial QSAR descriptor in this study was to investigate the possible effect of flavonoid lipophilicity upon their corresponding binding affinities.

Step 8. Steps 4−7 are repeated until all eleven, in this study, trial alignments are included in the 4D-QSAR analysis.

Step 9. The inspection and evaluation of the entire population of 4D-QSAR models is made in this step. The objective is to identify the "best" 4D-QSAR models with respect to the set of selected alignments. Each alignment leads to a particular best 4D-QSAR model for that specific alignment. The alignment corresponding to the 4D-QSAR model with the overall highest r^2 and xv-r^2 measures, for all alignments tested, is selected as the best alignment. The linear cross-correlation matrix of the GCODs for the best 4D-QSAR model for the optimal alignment is then built to determine the extent to which the GCODs of this best model are correlated to one another.

Step 10. The final step of the 4D-QSAR formalism is to hypothesize the "active" conformation of each flavonoid in the training set. This is achieved by first identifying all conformer states sampled for each compound, one at a time, that are within ΔE of the apparent global minimum energy conformation of the compound's CEP. Currently, ΔE is set at 2 kcal/mol. The resultant set of energy filtered conformations are then individually evaluated in the best 4D-QSAR model. That conformation within 2 kcal/mol of the apparent global minimum which predicts the highest activity in the best 4D-QSAR model is defined as the active conformation.

The 4D-QSAR model can also be used as a virtual high throughput screen, VHTS, to predict the biological activities of the members of a virtual library of "similar" compounds to those of a training set.[29,30] The use of a 4D-QSAR model as a VHTS is not novel, but rather an extended application of a fundamental purpose of constructing the QSAR model. QSAR models are constructed to permit forecasting of the biological activities of hypothetical molecules of interest. A comparison between the predicted activities from a 4D-QSAR model and the corresponding experimental values is

Table 4. Correlation Coefficient and Cross-Validated Correlation Coefficient of the Best 4D-QSAR Model for Each Alignment and Varying Number of Descriptors[a]

alignment	3 terms	4 terms	5 terms	6 terms	7 terms	8 terms
1	0.74	0.79	0.83	0.90	0.90	0.93
	0.71	0.76	0.80	0.84	0.84	0.90
2	0.68	0.83	0.86	0.91	0.95	0.95
	0.62	0.78	0.80	0.88	0.92	0.88
3	0.77	0.83	0.89	0.91	0.91	0.95
	0.72	0.78	0.85	0.87	0.87	0.92
4	0.76	0.82	0.87	0.91	0.92	0.94
	0.73	0.77	0.83	0.87	0.87	0.89
5	0.77	0.82	0.89	0.92	0.96	0.97
	0.73	0.77	0.80	0.88	0.90	0.93
6	**0.81**	**0.86**	**0.91**	**0.93**	**0.95**	**0.96**
	0.77	**0.83**	**0.87**	**0.90**	**0.92**	**0.92**
7	**0.77**	**0.86**	**0.90**	**0.91**	**0.94**	**0.95**
	0.74	**0.83**	**0.87**	**0.87**	**0.91**	**0.93**
8	0.74	0.86	0.89	0.92	0.93	0.93
	0.72	0.82	0.85	0.89	0.90	0.91
9	0.71	0.79	0.89	0.92	0.94	0.95
	0.65	0.75	0.84	0.89	0.88	0.88
10	0.80	0.82	0.89	0.91	0.94	0.95
	0.74	0.77	0.84	0.88	0.91	0.91
11	0.78	0.82	0.87	0.90	0.93	0.96
	0.75	0.77	0.84	0.86	0.90	0.94

[a]The correlation coefficient is the upper entry in each case. Correlation coefficients and cross-validated correlation coefficients of the best 4D-QSAR models for alignments 6 and 7 are shown in bold to emphasize that these two alignments give better models than the other nine alignments.

perhaps the most stringent validation test of any 4D-QSAR model. A virtual screen of a small library of four test flavonoids was carried out in this study.

RESULTS

Ten alignments (see Table 3) were chosen in order to span the core structure of flavone. They were divided into five sets focusing on the "left", "right", "left and middle", "right and middle", and the overall structure of the flavonoid as shown in Table 1 at the top. Models containing 3 terms to 8 terms for each alignment were generated. The r^2 and cross-validated r^2 (xv-r^2) of the corresponding best (RI) 4D-QSAR models are listed in Table 4. Alignments 6 and 7 spanning the "right" ring of the flavonoid structure yield the highest quality models. Based on these results, one more alignment, alignment 8 with three-ordered atoms chosen between those of alignments 6 and 7, was evaluated. However, the corresponding models of this alignment (see Table 4) do not have better features and fitting properties than those from alignments 6 and 7. The 4D-QSAR models from alignments 6 and 7 were, consequently, analyzed in detail and considered the most relevant models of the training set.

Alignment 6. Figure 1, Part A contains plots of r^2 and xv-r^2 as a function of the number of descriptors in the best corresponding model using alignment 6. The r^2 and xv-r^2 possess fairly high values, 0.81 and 0.77, respectively, when the model has only three descriptors, and these fitting measures continue to increase even when the model has eight descriptor terms. 4D-QSAR models having the same descriptor terms generally appear in different combinations over the top 10 models.

All unique 4D-QSAR models were determined for each set of 10 best models found for a given value of the GFA

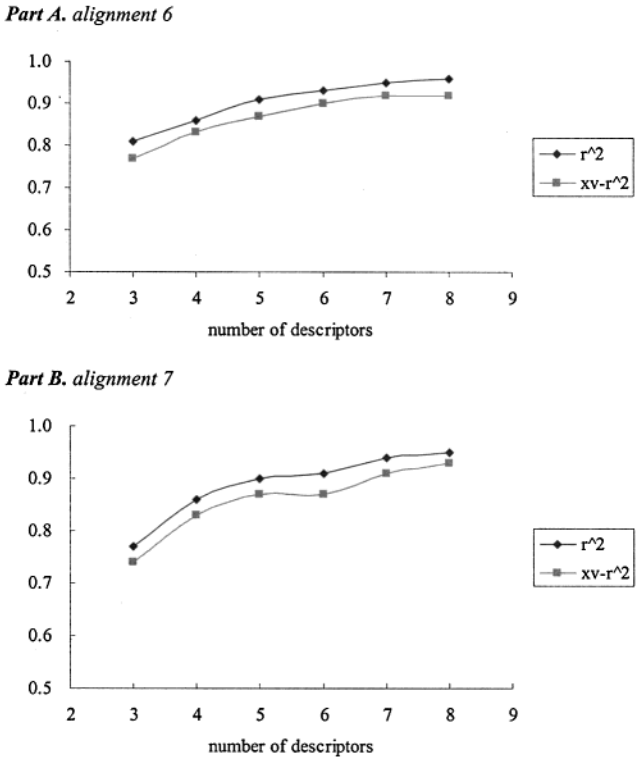


Figure 1. Plots of the number of descriptor terms present in the best 4D-QSAR models versus the r^2 and xv-r^2 of these models.

Table 5. Cross-Correlation Coefficient of the Residuals of Fit for the Best 4D-QSAR Models for Alignments Six and Seven

Part A. Alignment 6

model	3−1									
3−1[a]	1.00	3−2								
3−2	0.73	1.00	3−3							
3−3	0.75	0.76	1.00	5−1						
5−1	0.70	0.68	0.66	1.00	6−1					
6−1	0.65	0.62	0.61	**0.83**	1.00	7−1				
7−1	0.56	0.53	0.51	**0.74**	0.84	1.00	8−1			
8−1	0.54	0.56	0.55	**0.70**	0.79	0.79	1.00			

Part B. Alignment 7

model	3−1									
3−1	1.00	3−2								
3−2	0.71	1.00	4−1							
4−1	0.66	0.59	1.00	5−1						
5−1	0.83	0.74	**0.72**	1.00	5−2					
5−2	0.66	0.54	**0.65**	0.53	1.00	6−1				
6−1	0.64	0.64	**0.83**	0.67	0.76	1.00	6−2			
6−2	0.65	0.46	**0.81**	0.69	0.57	0.73	1.00	6−3		
6−3	0.50	0.29	0.44	0.42	0.38	0.44	0.70	1.00	7−1	
7−1	0.52	0.50	0.69	0.45	0.56	0.74	0.64	0.57	1.00	8−1
8−1	0.39	0.32	0.50	0.41	0.32	0.49	0.62	0.46	0.54	1.00
8−2	0.45	0.32	0.44	0.46	0.68	0.49	0.58	0.37	0.28	0.47

8−2
1.00

Part C. Alignment 7 with ClogP as an Additional Descriptor in the Trial Descriptor Pool

model	2−1									
2−1	1.00	2−2								
2−2	0.50	1.00	3−1							
3−1	0.76	0.65	1.00	3−2						
3−2	0.60	0.43	0.69	1.00	3−3					
3−3	0.72	0.40	0.54	0.64	1.00	3−4				
3−4	0.84	0.61	0.83	0.73	0.46	1.00	4−1			
4−1	0.51	0.44	0.61	0.81	0.60	0.52	1.00	4−2		
4−2	0.55	0.24	0.45	0.56	0.78	0.32	0.73	1.00	4−3	
4−3	0.53	0.75	0.67	0.62	0.54	0.66	0.49	0.37	1.00	5−1
5−1	0.46	0.64	0.68	0.66	0.65	0.55	**0.78**	**0.65**	**0.74**	1.00

[a] The number "3" means the model has three descriptor terms and the number "1" means the first model in a particular GFA optimization experiment, so "3−1" means the first 3 term model in the first set of optimizations.

smoothing factor. Unique models are defined as those models whose residuals of fit have a cross-correlation to the residuals of fit of at least one other top 10 model of less than 0.85. If two, or more, top 10 4D-QSAR models of a given GFA smoothing factor have a cross-correlation coefficient of their residuals of fit greater than or equal to 0.85, then the 4D-QSAR model with the fewest number of descriptor terms was selected to represent this common model family. The results of this analysis to identify all unique models are given in Table 5A. There is a high similarity in the location of the GCODs of all seven unique models listed. The MLR equations of the unique 4D-QSAR models of Table 5A are given in eqs 1−7. The coding "model (X−Y)" given above each MLR equation refers to the Yth most significant and unique 4D-QSAR model having X descriptor terms.

model 3−1

$$-\log Ki = -11.28*GC1(1,-6,0, any) + 14.89*GC2(1,3,6, p-) + 2.15*GC3(0,-7,0, any) + 6.67$$

$$n = 38 \quad r^2 = 0.83 \quad xv\text{-}r^2 = 0.78 \quad (1)$$

model 3−2

$$-\log Ki = -12.42*GC1(1,-6,0, any) + 14.14*GC2(1,3,6, p-) - 4.01*GC3(1,6,4, np) + 6.87$$

$$n = 38 \quad r^2 = 0.81 \quad xv\text{-}r^2 = 0.77 \quad (2)$$

model 3−3

$$-\log Ki = -12.02*GC1(1,-6,0, any) + 14.78*GC2(1,3,6, p-) - 6.09*GC3(1,2,6, p+) + 6.83$$

$$n = 38 \quad r^2 = 0.81 \quad xv\text{-}r^2 = 0.77 \quad (3)$$

model 5−1

$$\mathbf{-\log Ki = -12.38GC1(-1,-6,0, any) + 13.26GC2(-1,3,6, p-) - 7.56GC3(-1,6,5, any) + 1.76GC4(0,-7,0, p-) - 6.78GC5(1,1,6, p+) + 6.89}$$

$$\mathbf{n = 38 \quad r^2 = 0.89 \quad xv\text{-}r^2 = 0.84 \quad (4)}$$

model 6−1

$$-\log Ki = -14.10*GC1(1,-6,0, any) + 13.33*GC2(1,3,6, p-) - 8.74*GC3(-1,5,5, np) + 1.58*GC4(0,-6,-2, any) - 6.79*GC5(1,2,6, hbd) + 2.39*GC6(0,-3,3, any) + 6.26$$

$$n = 38 \quad r^2 = 0.93 \quad xv\text{-}r^2 = 0.90 \quad (5)$$

model 7−1

$$-\log Ki = -14.41*GC1(1,-6,0, any) + 12.46*GC2(1,3,6, p-) - 9.77*GC3(-1,5,5, np) + 1.81*GC4(0,-7,0, any) - 8.78*GC5(1,1,6, hbd) + 2.87*GC6(0,-3,3, any) + 0.62*GC7(0,0,4, np) + 5.83$$

$$n = 38 \quad r^2 = 0.95 \quad xv\text{-}r^2 = 0.92 \quad (6)$$

model 8−1

$$-\log Ki = -11.19*GC1(1,-6,0, any) + 9.98*GC2(1,3,6, p-) - 16.78*GC3(-1,5,6, np) + 3.27*GC4(0,-6,3, any) - 9.10*GC5(-1,2,6, p+) + 12.71*GC6(0,-3,-2, np) + 2.26*GC7(1,2,6, any) + 2.00*GC8(0,5,5, hba) + 5.83$$

$$n = 38 \quad r^2 = 0.94 \quad xv\text{-}r^2 = 0.91 \quad (7)$$

In eqs 1−7, GCi(x, y, z, T) is the ith significant GCOD located at (x, y, z) and having the T type IPE as defined in Table 2. The five GCODs in model 5−1 all appear among the GCODs of the three 3-term models. GC1(−1,−6,0, any) and GC2(−1,3,6, p-) in model 5−1 are each located very close to the GC1s and GC2s, respectively, in the 3-term models. GC3(−1,6,5, any) of model 5−1 is similar to GC3(1,6,4, np) in model 3−2. Both have negative correlation coefficients, indicating occupancy of this GCOD diminishes activity. GC4(0,−7,0, p-) of model 5−1 has the same location as GC3(0,−7,0, any) in model 3−1, but the IPEs are different. This difference of IPEs is probably the result of a lack of discriminating information about this pharmacophore position in the training set. GC5(1,1,6, p+) of model 5−1 is essentially in the same grid cell as GC3(1,2,6, p+) of model 3−3, since the distance between them is less than 1 Å. A pattern of GCOD similarity also is seen for all the other higher-term models (eqs 5−7).

The higher order term 4D-QSAR models are seemingly better able to describe the characteristics of the interaction between the flavonoids and the BzR site than the smaller (term) models, which is reflected in respective higher r^2 and $xv\text{-}r^2$ values. But, higher values of r^2 and $xv\text{-}r^2$ do not always mean that the corresponding quality of a model is also increasing. There is always the possibility of over-fitting of the data of the training set. If a model is over-fit, it may demonstrate a low residual of fit over the training set but not predict well. To avoid over-fitting, and to pick an optimal size 4D-QSAR model, the cross-correlation coefficient data in Table 5A were further investigated. The three unique 3-term models have moderate cross-correlation coefficients of their residuals of fit to one another but low cross-correlation coefficients of their residuals of fit to those of higher-term models. This observation suggests that the 3-term 4D-QSAR models do not provide as much and/or the same type of information about the interaction pharmacophore of the flavonoid with the BzR as the higher-term models. On the other hand, the residuals of fit of 4D-QSAR model 5−1 correlate fairly well with the residuals of fit of the higher-term 4D-QSAR models. The high cross-correlation coefficients are shown in bold in Table 5A and indicate that the 5-term 4D-QSAR model describes the ligand−receptor interaction nearly the same, and very nearly as well, as the higher-term models. Hence, 4D-QSAR model 5−1 was chosen as the best model under alignment 6 and eq 4 is singled out in bold print.

A stereographic representation of the 3D-pharmacophore embedded in 4D-QSAR model 5−1 is shown in Figure 2A. The reference compound in Figure 2A is compound 32, the most active compound in the training set (see Table 1), in its predicted active conformation. The observed and predicted −logKi values using eq 4 are plotted in Figure 3A. Two

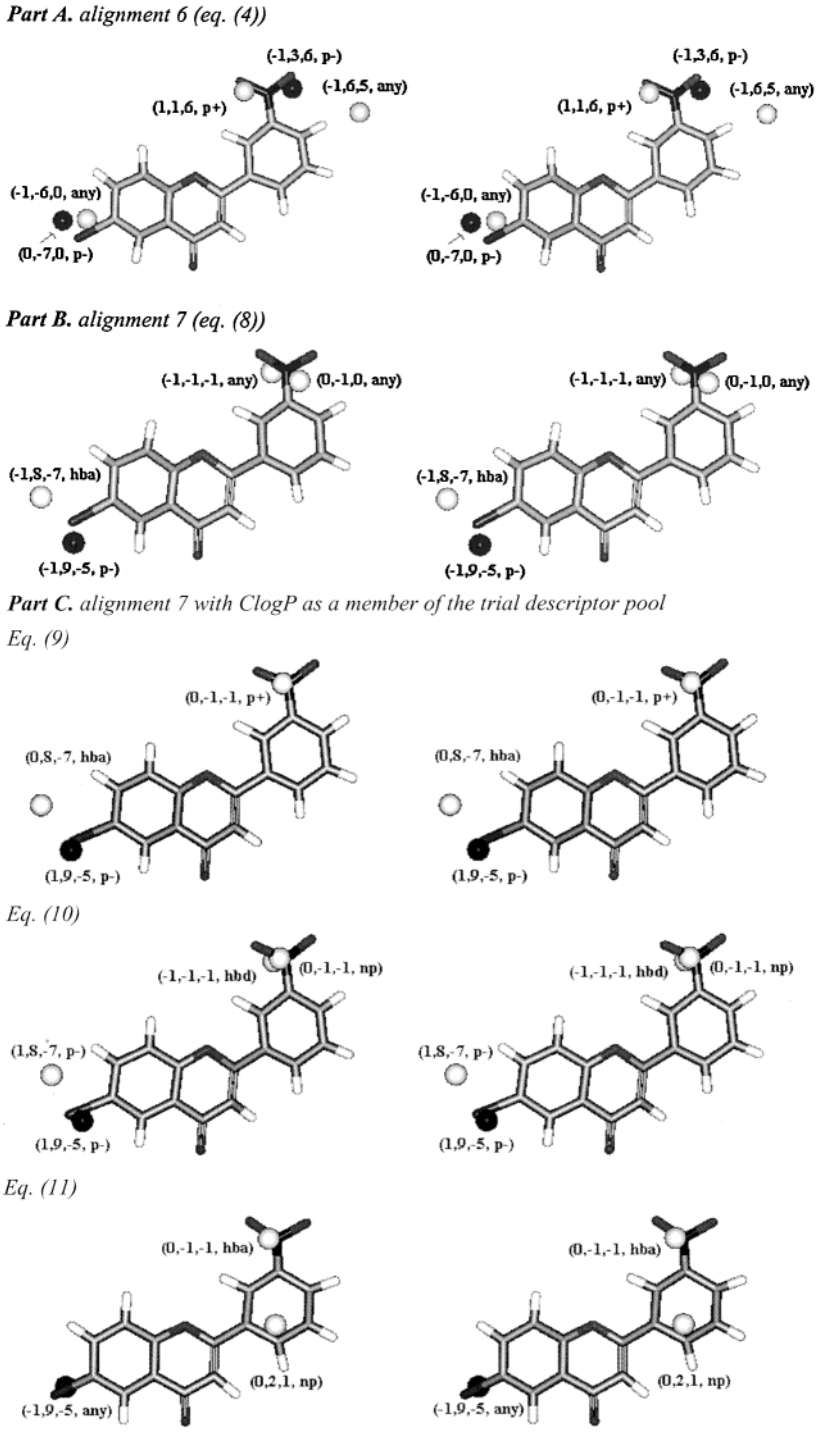


Figure 2. Stereoviews of the 3D-pharmacophores of the best 4D-QSAR models (eqs 4 and 8−11) shown relative to compound 32 (see Table 1), in its respective active conformations as predicted using each of the corresponding best 4D-QSAR models. The GCODs are shown as spheres, although the actual grid cells are cubes in space. The −logKi enhancing grid cells are shown as dark spheres, and grid cells which diminish −logKi are shown as light spheres. The spatial location and IPE type are listed next to each corresponding GCOD site.

GCODs of 4D-QSAR model 5−1 specify pharmacophore sites that enhance activity (i.e. these GCODs have positive regression coefficients) and correspond to

a) a polar negative IPE GCOD near $R_{3'}$, and

b) a polar negative IPE GCOD near the R_6 position.

Three GCODs of model 5−1 correspond to GCODs that decrease potency, since they have negative regression coefficients, and these 3D pharmacophore sites correspond to

c) two GCODs, basically having all-atom type IPEs, at the R_6 and $R_{4'}$ positions, and

d) a polar positive IPE GCOD near $R_{3'}$, which is consistent with the occupation of a polar negative GCOD near this position increasing receptor binding.

The best 4D-QSAR model, as represented by eq 4, has been statistically evaluated. A linear cross-correlation matrix of the GCODs of eq 4 was built and is given in Table 6A. Any pair of GCODs having a correlation of greater than 0.50, or less than −0.50, is flagged as a highly correlated pair. For alignment 6, none of the GCODs are significantly correlated to one another.

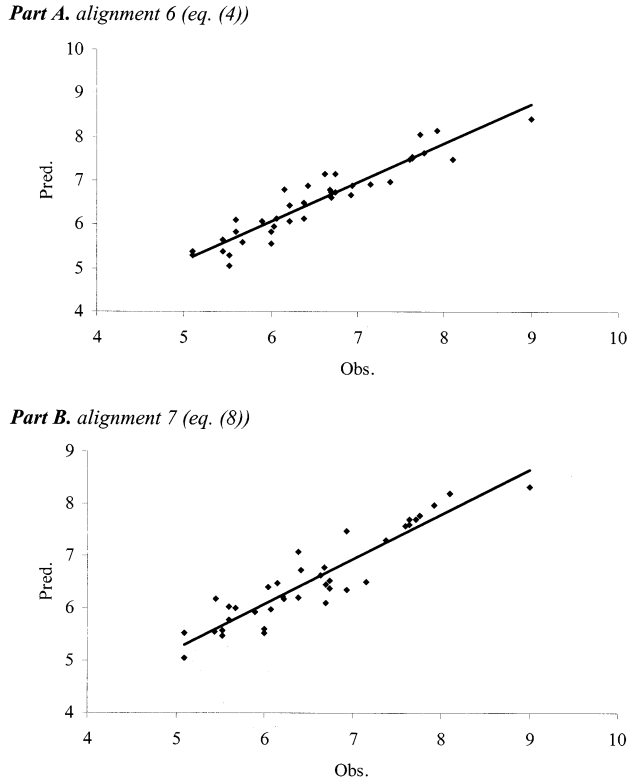


Figure 3. Predicted and observed activities (−logKi) of flavonoid analogues for alignments 6 and 7 (and eqs 4 and 8), respectively.

Activity profiling of a 4D-QSAR model involves generating a matrix containing the contribution of each GCOD of each compound in the training set to the predicted activity of the compound. The activity-profile matrix of eq 4 is presented in Table 7A. The activity-profiling matrix also illustrates the relatively distinct fingerprint made by the set of GCODs of each compound to the −logKi value of that compound. All the compounds occupy GC1. GC2 is occupied by compounds that are substituted at C14. Compounds 13 to 16, having −F at the C14 position, have less contribution from GC2 than compounds 21 to 24 and 25 to 28 which have the bulkier groups, −Cl and −Br, respectively. It appears that GC3 is present in the 4D-QSAR model largely to accommodate analogues 7, 8 and 30. These three compounds have bulkier groups at $R_{4'}$ than the other training set compounds, and the activities (i.e. [−logKi]) of compound 7, 8 and 30 are corresponding lower than the activities of analogues 19 and 20 which have the identical structures except for groups at the $R_{4'}$ position. The contribution from GC4 to −logKi is quite low, except for compounds 10 and 31. These two compounds have a $-NO_2$ group at R_6, while the rest of the compounds have smaller groups occupying GC4. GC5 is present mainly to capture the influence of the −OH at $R_{3'}$ for analogues 2, 3, and 4 on −logKi.

ClogP was included in the trial descriptor pool in the construction of eq 1 but did not survive in GFA model optimization. The absence of ClogP as a descriptor in the best 4D-QSAR model suggests that total molecular lipophilicity is not a significant activity determinant. Moreover, the role of lipophilicity is seemingly captured by the appropriate IPE type of GCODs.

Table 6. Linear Cross-Correlation Matrices of the GCOD Descriptors Found in the Best 4D-QSAR Models

	(−1,−6,0,any)	(−1,3,6,p-)	(−1,6,5,any)	(0,−7,0,p-)	(1,1,6,p+)
		Part A. Alignment 6 (Eq 4)			
(−1,−6,0,any)	1.00				
(−1,3,6,p-)	−0.05	1.00			
(−1,6,5,any)	−0.21	−0.17	1.00		
(0,−7,0,p-)	−0.1	0.06	−0.05	1.00	
(1,1,6,p+)	−0.15	−0.09	−0.06	−0.06	1.00
	(−1,−1,−1,any)	(−1,8,−7,hba)	(−1,9,−5,p-)	(0,−1,0,any)	
		Part B. Alignment 7 (Eq 8)			
(−1,−1,−1,any)	1.00				
(−1,8,−7,hba)	−0.07	1.00			
(−1,9,−5,p-)	0.18	**0.61**	1.00		
(0,−1,0,any)	−0.25	−0.07	−0.14	1.00	
	(0,−1,−1,p+)	(0,8,−7,hba)	(1,9,−5,p-)	logP	
		Part C. Alignment 7 with ClogP as a Member of the Trial Descriptor Pool Eq 9			
(0,−1,−1,p+)	1.00				
(0,8,−7,hba)	−0.06	1.00			
(1,9,−5,p-)	0.24	**0.58**	1.00		
logP	−0.12	0.15	0.39	1.00	
	(−1,−1,−1,hbd)	(0,−1,−1,np)	(1,8,−7,p-)	(1,9,−5,p-)	
		Eq 10			
(−1,−1,−1,hbd)	1.00				
(0,−1,−1,np)	−0.26	1.00			
(1,8,−7,p-)	−0.08	0.20	1.00		
(1,9,−5,p-)	0.24	−0.22	**0.60**	1.00	
	(−1,9,−5,any)	(0,−1,−1,p-)	(0,2,1,np)	logP	
		Eq 11			
(−1,9,−5,any)	1.00				
(0,−1,−1,p-)	0.10	1.00			
(0,2,1,np)	−0.24	−0.30	1.00		
logP	0.30	−0.13	−0.02	1.00	

Table 7. Activity-Profile Matrices of the GCODs of the Best 4D-QSAR Models[a]

compd	−12.38*GC1 (−1,−6,0,any)	13.26*GC2 (−1,3,6,p-)	−7.56*GC3 (−1,6,5,any)	1.76*GC4 (0,−7,0,p-)	−6.78*GC5 (1,1,6,p+)	compd	−12.38*GC1 (−1,−6,0,any)	13.26*GC2 (−1,3,6,p-)	−7.56*GC3 (−1,6,5,any)	1.76*GC4 (0,−7,0,p-)	−6.78*GC5 (1,1,6,p+)
					Part A. Alignment 6 (Eq 4)						
1	−1.07	0.00	0.00	0.00	0.00	20	−0.06	0.00	0.00	0.04	0.00
2	−1.03	0.26	0.00	0.00	−0.27	21	−1.34	0.89	0.00	0.00	0.00
3	−0.33	0.33	0.00	0.00	−0.75	22	−1.02	0.76	0.00	0.00	0.00
4	−0.10	0.08	0.00	0.03	−0.82	23	−0.37	1.02	0.00	0.00	0.00
5	−0.69	0.95	0.00	0.00	0.00	24	−0.14	0.88	0.00	0.01	0.00
6	−0.15	0.74	0.00	0.00	0.00	25	−1.66	0.89	0.00	0.00	0.00
7	−0.18	0.00	−0.65	0.00	0.00	26	−1.14	1.39	0.00	0.00	0.00
8	−0.12	0.00	−1.20	0.02	0.00	27	−0.26	0.88	0.00	0.01	0.00
9	−0.12	0.00	0.00	0.02	0.00	28	−0.11	1.25	0.00	0.03	0.00
10	−0.41	0.00	0.00	1.00	0.00	29	−0.04	0.00	0.00	0.06	0.00
11	−0.41	0.00	0.00	0.01	0.00	30	−0.11	0.00	−0.08	0.04	0.00
12	−0.02	0.00	0.00	0.02	0.00	31	−0.72	1.38	0.00	0.59	0.00
13	−1.42	0.18	0.00	0.00	0.00	32	−0.09	1.58	0.00	0.03	0.00
14	−1.06	0.12	0.00	0.00	0.00	33	−0.11	0.00	0.00	0.02	0.00
15	−0.31	0.10	0.00	0.00	0.00	34	−1.84	0.00	0.00	0.00	0.00
16	−0.14	0.20	0.00	0.02	0.00	35	−1.59	0.00	0.00	0.00	0.00
17	−1.52	0.00	0.00	0.00	0.00	36	−1.52	0.00	0.00	0.00	0.00
18	−0.79	0.00	0.00	0.00	0.00	37	−1.61	0.00	0.00	0.00	0.00
19	−0.16	0.00	0.00	0.01	0.00	38	−1.33	0.00	0.00	0.00	0.00

compd	−18.79*GC1 (−1,−1,−1, any)	−11.36*GC2 (−1,8,−7, hba)	18.36*GC3 (−1,9,−5, p-)	−2.00*GC4 (0,−1,0, any)	compd	−18.79*GC1 (−1,−1,−1, any)	−11.36*GC2 (−1,8,−7, hba)	18.36*GC3 (−1,9,−5, p-)	−2.00*GC4 (0,−1,0, any)
				Part B. Alignment 7 (Eq 8)					
1	0.00	0.00	0.00	−0.62	20	0.00	0.00	0.78	−0.64
2	−0.64	0.00	0.43	0.00	21	0.00	0.00	0.00	0.00
3	−1.84	0.00	1.60	−0.01	22	0.00	0.00	0.22	0.00
4	−1.74	0.00	1.70	−0.01	23	0.00	0.00	1.37	0.00
5	0.00	0.00	0.30	0.00	24	0.00	0.00	1.56	0.00
6	0.00	0.00	1.97	0.00	25	−0.02	0.00	0.00	0.00
7	0.00	0.00	0.39	−0.69	26	−0.10	0.00	0.50	0.00
8	0.00	0.00	0.46	−0.67	27	−0.05	0.00	1.51	0.00
9	0.00	0.00	2.26	−1.70	28	−0.04	0.00	1.52	0.00
10	0.00	−1.69	3.44	−0.40	29	0.00	0.00	0.87	−0.60
11	0.00	0.00	1.59	−0.74	30	0.00	0.00	0.22	−0.35
12	0.00	0.00	1.20	−0.70	31	0.00	−1.10	2.85	0.00
13	0.00	0.00	0.00	−0.05	32	0.00	0.00	2.10	0.00
14	0.00	0.00	0.18	−0.01	33	0.00	0.00	0.85	−0.59
15	0.00	0.00	1.29	−0.04	34	0.00	0.00	0.00	−0.65
16	0.00	0.00	1.11	−0.04	35	0.00	0.00	0.00	−0.74
17	0.00	0.00	0.00	−0.66	36	0.00	0.00	0.00	−1.17
18	0.00	0.00	0.20	−0.65	37	0.00	0.00	0.00	−0.71
19	0.00	0.00	0.82	−0.66	38	0.00	0.00	0.03	−0.71

compd	−8.49*GC1 (0,−1,−1, p+)	−5.12*GC2 (0,8,−7, hba)	18.16*GC3 (1,9,−5, p-)	0.36* ClogP	compd	−8.49*GC1 (0,−1,−1, p+)	−5.12*GC2 (0,8,−7, hba)	18.16*GC3 (1,9,−5, p-)	0.36* ClogP
				Part C. Alignment 7 with ClogP as a Member of the Trial Descriptor Pool Eq 9					
1	0.00	0.00	0.00	0.84	20	0.00	0.00	0.79	1.17
2	−0.49	0.00	0.25	0.78	21	0.00	0.00	0.00	1.02
3	−1.37	0.00	1.72	0.92	22	0.00	0.00	0.24	1.07
4	−1.43	0.00	1.85	1.02	23	0.00	0.00	1.26	1.21
5	0.00	0.00	0.36	0.87	24	0.00	0.00	1.55	1.31
6	0.00	0.00	1.94	1.00	25	0.00	0.00	0.00	1.17
7	0.00	0.00	0.55	0.93	26	0.00	0.00	0.48	1.17
8	0.00	0.00	0.50	1.03	27	0.00	0.00	1.45	1.31
9	0.00	0.00	0.38	1.10	28	0.00	0.00	1.46	1.40
10	0.00	−1.85	3.29	1.10	29	0.00	0.00	0.86	1.12
11	0.00	0.00	1.55	1.07	30	0.00	0.00	0.27	1.10
12	0.00	0.00	1.12	1.17	31	0.00	−0.58	2.78	0.80
13	0.00	0.00	0.00	0.89	32	0.00	0.00	2.35	1.10
14	0.00	0.00	0.15	0.94	33	0.00	0.00	0.75	1.20
15	0.00	0.00	1.15	1.07	34	0.00	0.00	0.00	0.63
16	0.00	0.00	1.06	1.17	35	0.00	0.00	0.00	0.53
17	0.00	0.00	0.00	0.89	36	0.00	0.00	0.00	0.82
18	0.00	0.00	0.11	0.94	37	0.00	0.00	0.00	0.68
19	0.00	0.00	0.74	1.07	38	0.00	0.00	0.03	0.44

Table 7 (Continued)

compd	−19.27*GC1 (−1,−1,−1, hbd)	−0.68*GC2 (0,−1,−1, np)	−9.31*GC3 (1,8,−7, p-)	18.95*GC4 (1,9,−5, p-)	compd	−19.27*GC1 (−1,−1,−1, hbd)	−0.68*GC2 (0,−1,−1, np)	−9.31*GC3 (1,8,−7, p-)	18.95*GC4 (1,9,−5, p-)
				Eq 10					
1	0.00	−0.47	0.00	0.00	20	0.00	−0.46	−0.04	0.82
2	−0.66	0.00	0.00	0.27	21	0.00	0.00	0.00	0.00
3	−1.89	0.00	0.00	1.79	22	0.00	0.00	0.00	0.25
4	−1.78	0.00	−0.01	1.93	23	0.00	0.00	−0.01	1.32
5	0.00	0.00	0.00	0.38	24	0.00	0.00	−0.01	1.62
6	0.00	0.00	0.00	2.03	25	0.00	0.00	0.00	0.00
7	0.00	−0.45	−0.05	0.58	26	0.00	0.00	0.00	0.50
8	0.00	−0.45	−0.21	0.52	27	0.00	0.00	−0.01	1.52
9	0.00	−0.10	0.00	0.40	28	0.00	0.00	−0.04	1.53
10	0.00	−0.54	−1.36	3.43	29	0.00	−0.48	−0.07	0.90
11	0.00	−0.43	0.00	1.62	30	0.00	−0.56	−0.25	0.28
12	0.00	−0.44	−0.05	1.17	31	0.00	0.00	−0.79	2.90
13	0.00	0.00	0.00	0.00	32	0.00	0.00	0.00	2.45
14	0.00	0.00	0.00	0.15	33	0.00	−0.48	−0.02	0.79
15	0.00	0.00	−0.01	1.20	34	0.00	−0.46	0.00	0.00
16	0.00	0.00	−0.05	1.11	35	0.00	−0.43	0.00	0.00
17	0.00	−0.46	0.00	0.00	36	0.00	−0.28	0.00	0.00
18	0.00	−0.46	0.00	0.11	37	0.00	−0.44	0.00	0.00
19	0.00	−0.46	−0.01	0.77	38	0.00	−0.44	0.00	0.03

compd	−1.09*GC1 (0,−1,−1, hba)	−4.13*GC2 (0,2,1, np)	8.71*GC3 (−1,9,−5, any)	0.59* ClogP	compd	−1.09*GC1 (0,−1,−1, hba)	−4.13*GC2 (0,2,1, np)	8.71*GC3 (−1,9,−5, any)	0.59* ClogP
				Eq 11					
1	0.00	−3.44	0.00	1.37	20	0.00	−3.69	0.37	1.92
2	−1.09	−2.95	0.20	1.28	21	0.00	−3.44	0.01	1.68
3	−1.08	−2.86	0.76	1.50	22	0.00	−3.44	0.10	1.76
4	−1.09	−2.98	0.81	1.67	23	0.00	−3.39	0.65	1.98
5	0.00	−2.33	0.14	1.42	24	0.00	−3.36	0.74	2.14
6	0.00	−2.32	0.94	1.65	25	0.00	−3.39	0.01	1.92
7	0.00	−3.49	0.18	1.52	26	0.00	−3.18	0.24	1.92
8	0.00	−3.60	0.22	1.69	27	0.00	−3.39	0.72	2.14
9	0.00	−3.94	1.07	1.81	28	0.00	−3.32	0.72	2.30
10	0.00	−3.76	1.63	1.81	29	0.00	−3.51	0.41	1.83
11	0.00	−3.78	0.75	1.76	30	0.00	−3.29	0.10	1.81
12	0.00	−3.75	0.57	1.92	31	0.00	−2.36	1.35	1.32
13	0.00	−3.36	0.01	1.45	32	0.00	−2.33	1.00	1.81
14	0.00	−3.63	0.09	1.53	33	0.00	−3.68	0.41	1.96
15	0.00	−3.39	0.61	1.76	34	0.00	−3.50	0.00	1.03
16	0.00	−3.40	0.53	1.92	35	0.00	−3.46	0.00	0.86
17	0.00	−3.73	0.00	1.45	36	0.00	−3.86	0.03	1.34
18	0.00	−3.65	0.10	1.53	37	0.00	−3.75	0.01	1.12
19	0.00	−3.77	0.39	1.76	38	0.00	−3.75	1.10	0.71

[a] Each column after the first column contains the contribution to −logKi of each term in a 4D-QSAR model for each compound in the training set.

Alignment 7. The best values of r^2 and xv-r^2 for the number of descriptors in 4D-QSAR models using alignment 7 are shown in Figure 1B. The 3-term models have good quality (see Table 4), and the quality is still improving even when the number of descriptors reaches eight. The same procedure to extract the best 4D-QSAR model for alignment 6 was repeated for alignment 7, see Table 5B. All eleven low cross-correlated 4D-QSAR models listed in Table 5B share similar GCODs. Model 4-1 correlates fairly well with the five, six and seven term 4D-QSAR models. These high cross-correlation coefficients are shown in bold in Table 5B. However, model 4−1 correlates poorly with models 8−1 and 8−2, and the five, six and seven term 4D-QSAR models also correlate poorly with models 8−1 and 8−2. Hence, model 4−1 is a seemingly representative and concise 4D-QSAR model for all models of alignment 7 and has been chosen as the best 4D-QSAR model under alignment 7.

$$-\log Ki = 6.22 - 18.79GC1(-1,-1,-1, \text{any}) - 11.36GC2(-1,8,-7, \text{hba}) + 18.36GC3(-1,9,-5, \text{p-}) - 2.00GC4(0,-1,0, \text{any})$$

$$n = 38 \qquad r^2 = 0.86 \qquad \text{xv-}r^2 = 0.83 \qquad (8)$$

The stereographic representation of the 3D-pharmacophore of the 4D-QSAR model, eq 8, is shown in Figure 2B. The reference compound is again compound 32 (see Table 1), in its predicted active conformation based on eq 8. The observed and predicted −logKi values, using eq 8, are plotted in Figure 3B. One GCOD of eq 8 has a positive regression coefficient (i.e. its occupancy enhances activity) and corresponds to

a) a polar negative IPE GCOD near the R_6 position.

The other three GCODs of eq 8 have negative regression coefficients and correspond to

b) a hydrogen bond acceptor site near R_6, and

c) two GCODs with all-atom type IPEs characterizing the steric restriction associated with the $R_{3'}$ position.

The linear cross-correlation matrix of the GCODs of eq 8 is given in Table 6B. GC2 and GC3 have a cross-correlation coefficient of 0.61, indicating they are significantly correlated to one another. Removal of GC3 from the descriptor pool, followed by re-optimization of the 4D-QSAR models yields an "optimum" 3-term 4D-QSAR model that is statistically poor ($r^2 < 0.60$). However, the best 4D-QSAR model generated after removal of GC2 from the trial descriptor pool has a r^2 of 0.79. This model optimization behavior is analyzed below in terms of the activity profile matrix.

The activity profile matrix of each GCOD of the best 4D-QSAR model using alignment 7, eq 8, is given in Table 7B. GC1, which is similar to GC5 in eq 4, accommodates compounds 2, 3 and 4 of the training set, as well as capturing information about compounds 25 to 28, which all have −Br at C14. Other analogues having smaller substituents are not able to reach and occupy this grid cell. Two compounds, 10 and 31, largely occupy GC2, which is similar to GC4 of eq 4. GC3, which is a polar negative GCOD, dominates the activity profile matrix in terms of descriptor magnitude and also in terms of variability across the members of the training set. Only compounds that have substituents at C3 are able to occupy GC3. Interestingly, GC3 contributes the most to specifying the −logKi values of compounds 10 and 31, the only two compounds that also have contributions to −logKi from GC2. This joint dependence of −logKi on both GC2 and GC3 may explain the high cross-correlation of GC2 and GC3. Since only two analogues from the set of 38 occupy GC2, the removal of GC2 does not decrease r^2 significantly but is crucial to the successful prediction of compounds 10 and 31 and, presumably, other possible analogues having the same substituent pattern.

ClogP was again included in the trial descriptor pool and did survive in the GFA 4D-QSAR model optimization process. Thus, alignment 7 models with ClogP were further explored and are described below.

Alignment 7 with ClogP as an Additional Descriptor. A set of optimized 4D-QSAR models from two to five descriptor terms were generated in which ClogP often survived as a descriptor in a final model. The cross-correlation coefficients among these models are listed in Table 5C. The higher-term models generally include all the GCODs appearing in lower-term models and the less significant GCODs permitted by the larger allowed model size. Three 4-term models (4−1, 4−2, 4−3 of Table 5C) are dissimilar from one another as judged by the low cross-correlation coefficients of their residuals of fit (see Table 5C). Model 4-3 is particularly distinct from the other two 4-term models. The cross-correlation coefficients of the residuals of fit of these three 4-term models with 5-term model 5−1 are shown in bold in Table 5C. Although the three models are different from one another, they correlate with model 5−1 rather well. This observation suggests that the 4-term models are able to capture much of the information of the best 5-term model but in different ways. To avoid the over-fitting phenomena, and in order to make a comparison with the 4-term model from alignment 7 (i.e. eq 2), all three 4-term models have been selected.

model 4−1

$$-\log Ki = -8.49GC1(0,-1,-1, p+) - 5.12GC2(0,8,-7, hba) + 18.16GC3(1,9,-5, p-) + 0.36ClogP + 5.05$$

$$n = 38 \quad r^2 = 0.82 \quad xv\text{-}r^2 = 0.78 \qquad (9)$$

model 4−2

$$-\log Ki = -19.27GC1(-1,-1,-1, hbd) - 0.68GC2(0,-1,-1, np) - 9.31GC3(1,8,-7, p-) + 18.95GC4(1,9,-5, p-) - 6.11$$

$$n = 38 \quad r^2 = 0.81 \quad xv\text{-}r^2 = 0.76 \qquad (10)$$

model 4−3

$$-\log Ki = 8.71*GC1(-1,9,-5, any) - 1.09*GC2(0,-1,-1, hba) - 4.13*GC3(0,2,1, np) + 0.59ClogP + 8.20$$

$$n = 38 \quad r^2 = 0.86 \quad xv\text{-}r^2 = 0.82 \qquad (11)$$

The stereographic representations of the 3D-pharmacophores embedded in eqs 9−11 are shown in Figure 2C from the same external reference view as in Figure 2A,B. The reference analogue is again analogue 32, but in the active conformations predicted by each of the three 4-term 4D-QSAR models (eqs 9−11). All three models have a grid cell near R_6 and one near $R_{3'}$. Model 4-1 and 4-2 share one grid cell site between R_6 and R_7. Model 4-3 identifies a sterically restricted site near C17 (see Table 1).

The linear cross-correlation matrices of the GCODs of eqs 9−11 are given in Table 6C. In model 4−1, GC2(0,8,−7, hba) and GC3(1,9,−5, p-) have a cross-correlation coefficient of 0.58. Two GCODs in model 4−2 (GC3 (1,8,−7, p-) and GC4(1,9,−5, p-)) are close in space to GC2 and GC3 of model 4−1. GC3 and GC4 of 4D-QSAR model 4−2 have a cross-correlation coefficient greater than 0.50. An analysis of the linear cross-correlation matrices of the GCODs of eqs 9 and 11 indicates a GCOD descriptor, which has a polar negative IPE, or a hydrogen bond acceptor, near (0,8,−7) is somehow associated with the GCOD descriptor having a polar negative IPE near (1,9,−5). The former GCOD descriptor contributes negatively to the activity, while the latter has a positive regression coefficient. This behavior of the GCODs suggests that −logKi is quite sensitive to substituents at R_6 with respect to electrostatic character and steric size. The activity profile matrices for the 4D-QSAR models represented by eqs 9−11 are given in Table 7C. The ClogP values are reported in the profile matrix, but they cannot be compared in a meaningful way to the GCOD descriptor values.

Finally, 2-term 4D-QSAR models (i.e. 2−1 and 2−2) were generated by increasing the GFA smoothing factor[14] with the express goal to see which descriptors might remain in a very small 4D-QSAR model. The descriptors of a small model may be the most basic and important features of a ligand−receptor interaction. GC(−1,9,−5, p-),

Table 8. Predicted and Observed Activities (−logKi) for the Test Set Using the Best Models from Alignments Six and Seven, Alignment Seven with ClogP as a Descriptor

		pred.				
	obs.	eq 4	eq 8	eq 9	eq 10	eq 11
baicalin	4.11	5.80	5.47	5.11	5.65	**4.91**
baicalein	5.25	5.89	5.69	**5.65**	5.73	6.82
scutellarein	4.92	5.92	5.58	**5.55**	5.76	6.49
wogonin	5.69	5.32	**5.63**	5.59	**5.63**	5.37

Table 9. Grid Cell Occupancy Values of the GCODs under Alignment Seven for Four Test Compounds

compound	GC (−1,9,−5,any)	GC (−1,9,−5, p-)	GC (1,9,−5, any)	GC (1,9,−5, p-)
baicalin	0.017	0.008	0.014	0.004
baicalein	0.154	0.008	0.164	0.004
scuttelarein	0.136	0.007	0.123	0.004
wogonin	0.000	0.000	0.000	0.000

GC(0,−1,−2, hba) and logP remain as descriptors among the 2-term models. This finding further suggests that substituents near R_6 and $R_{3'}$ are most crucial to ligand−receptor binding for this training set.

Virtual Screening. The last four compounds (drugs) in Table 1 have been used as the test set for virtual screening using equations (4, 8, 9−11). The experimental and the predicted −logKi values are listed in Table 8. These four compounds have been used as a test in a previous study[10] which is the main reason they have been adopted as a virtual test set in this work. However, there four compounds are, in composite, rather inactive relative to the training set and, thus, largely explore how well a QSAR model can predict inactivity.

The best prediction for each test compound is indicated in bold. The test compounds have lower −logKi values as compared to most compounds in the training set. The binding of wogonin is predicted well by every 4D-QSAR model. Wogonin is very similar in chemical structure to the training set compounds, which would suggest that its −logKi should be well predicted by the 4D-QSAR models of the training set. Baicalin has a very bulky and hydrophilic substituent group at R_7. None of the compounds in the training set possess this structural feature. Still eqs 9 and 11 include the ClogP descriptor which captures the hydrophilic character of the 5-carboxyglucosyloxy group. Hence, the activity of baicalin is best predicted by eqs 9 and 11. Baicalein and scutellarein each have three hydroxyl groups at R_5, R_6 and R_7. Intramolecular hydrogen bond formation among these hydroxyls may possibly influence the electron density of the oxygen atom at R_6. Equation 11 gives the worst predictions of the −logKi values for baicalein and scutellarein. A study of the differences between eq 11 and the other 4D-QSAR models indicates that the IPE types for the GCODs at (−1,9,−5) or (1,9,−5) are crucial and might be the cause of the poor predictions using eq 11. Table 9 lists the occupancies of the test compounds at the same grid cell, but using different IPE types. The four test compounds have lower ClogP values than almost every compound in the training set. The correlation coefficients between the values of −logKi and ClogP increase from 0.60 to 0.71 after including the four test compounds. This explains that eq 9 having ClogP as one descriptor, generally, gives better predictions of test compounds than the other models.

Table 10. Cross-Correlations of the Residuals of Fit for the Best 4D-QSAR Models Generated from the 38-Compound Training Set

	eq 4	eq 8	eq 9	eq 10	eq 11
eq 4	1.00				
eq 8	0.36	1.00			
eq 9	0.49	0.61	1.00		
eq 10	0.31	**0.85**	0.73	1.00	
eq 11	0.52	0.32	0.49	0.37	1.00

The results of the virtual screening suggest that there is more binding specificity from structural features of the flavone analogues than can be extracted from the 4D-QSAR models developed for the training set. To better understand the ligand−receptor interaction, the four test compounds were added to the training set to develop 4D-QSAR models for all 42 compounds. Alignment 6 and 7 were applied to the expanded 42-compound training set. The same model optimization and exploration analyses used for the original 38-compound training set was applied to the 42-compound training set. However, the addition of the four test compounds to the 38-compound training set does not meaningfully alter, or add new features to, the forms of the best 4D-QSAR models but rather "refines" the descriptor terms of the models.

DISCUSSION

There are significant similarities between the five 4D-QSAR models generated from the 38-compound training set. As shown in Figure 2, the GCODs are focused on the $R_{3'}$ and R_6 positions. The −logKi values of the compounds are highly dependent on the size and electrostatics character of the substituents at these two positions based on the IPEs of the GCODs near these two substituent sites. Polar negative groups correctly embedded in the $R_{3'}$ and/or R_6 substitutents are predicted to increase −logKi values as can be gleaned from the p- IPE of a GCOD near each substituent site.

An investigation of eq 4 indicates that the order of the substituents in the training set at R_6, in terms of the magnitude of their negative GC1 occupancy contributions to −logKi, is, $-OH > -H > -F > -NO_2 > -Cl > -Br$ (see Table 7A). Interestingly, the bulkier groups, $-NO_2$ and −Br, are predicted to balance their negative GC1 contribution to −logKi by also occupying the positive contributing descriptor to −logKi, GC4(0,−7,0, p-), while the smaller R_6 substituents cannot "reach" this GCOD. For the major positive descriptor to −logKi, GC2(−1,3,6, p-), the order of the substituents contribution to −logKi is, $-H < -F \approx -OH < -Cl \approx -NO_2 \approx -Br$. The most active compound in the training set, compound 32, has a −Br at R_6 and a $-NO_2$ at $R_{3'}$, while two of the most inactive compounds, 36 and 37, have the unfavored −OH at R_6 and no substituent at $R_{3'}$.

Equation 8 has the same overall 3D-pharmacophore profile as eq 4. Compounds 36 and 37 do not occupy the enhancing grid cell (−1, 9, −5) with a p- IPE, whereas compound 32 has significant occupation of this grid cell by a p- IPE.

Although the top five 4D-QSAR models, eqs (4, 8, 9−11), generated from the 38-compound training set correspond to similar 3D pharmacophores, these models are not highly correlated to one another, with the exception of eqs 8 and 10. The cross-correlation coefficients of the residuals of fit for these five 4D-QSAR models are listed in Table 10. Overall, this finding suggests that each of the five models is

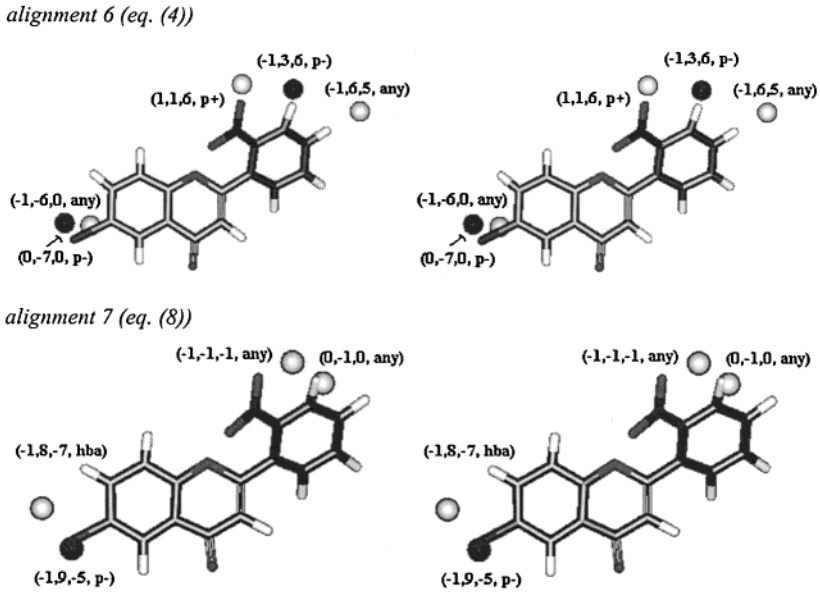


Figure 4. Stereoviews of the active conformations of compound 9 predicted by eqs 4 and 8, respectively.

capturing unique "pieces" of information about receptor binding. For example, eq 4 captures binding features for substituents at $R_{4'}$, while eq 8 does not provide this information. Therefore, the composite "picture" of the interaction of flavonoids and the BzR binding site should be conceptualized from all five models. That is, the five best 4D-QSAR models should be considered as a single consensus model.

Five compounds (9, 11, 12, 36 and 37) of the training set have substituents at position $R_{2'}$, while the rest of the training set compounds are unsubstituted at $R_{2'}$. $R_{2'}$ is close to the oxygen at position 7 (see Table 1). Hence, it is possible to speculate that the active conformations of these five compounds will not be planar due to steric crowding. The stereographic representation of compound 9 in its predicted active conformations using eqs 4 and 8, which are very similar nonplanar conformations, are shown in Figure 4. The torsion angle between the rings for both compounds 36 and 37 is almost 30 degrees (i.e. right ring C is tilted from the plane of left ring A and middle ring B.), so that the predicted active conformations of these compounds are similar to those of compound 9. The active conformations of flavone (i.e. compound 1) as predicted by eqs 4 and 8 also are not strictly planar but slightly bow-shaped. The left ring A and right ring C tilt up from middle ring B with a curvature of approximately 10 degrees. However, compounds 11 and 12 have predicted active conformations close to being planar based on eqs 4 and 8 but energetically could readily adopt conformations such as that of the active conformation of compound 9.

The predicted active conformations of the remainder of the compounds in the training set are also bow-shaped like those of flavone or "twisted" like the active conformation of compound 9. The "twisting" in compound 9 arises from ring C being tilted away from the plane of rings A and B. Moreover, some compounds have active conformations that are combinations of the "plane" of rings A and B being slightly bow-shaped and ring C tilted from this plane. In other words, the chemical structure of each compound controls the curvative of the A and B rings and the rotamer state of the torsion angle between the B and C rings.

The carbonyl oxygen plays a crucial role in deciding the active conformations of the flavonoids.[31,32] Stabilization of the partial negative charge on the oxygen atom increases the contribution of the extensive mesomeric effect, thus, favoring the planar conformation. There is a delicate balance between the mesomeric effect and the repulsive steric interaction. Presumably, when the flavonoids bind with the $GABA_A$ receptor, they seek the optimum conformation to maximally occupy the activity-enhancing GCODs and minimally occupy the activity-decreasing GCODs under this balance. The torsion energy required to rotate about the single bond between rings B and C is usually low for compounds in the training set that do not have adjacent bulky groups.[8,31−33] For these types of analogues, the energy balance tends to favor the planar conformation, such as is seen for compound 32. Therefore, it appears that although the overall structures of the compounds in the training set are relatively rigid, each compound take full advantage of its individual flexibility to reach its optimum active conformation, which, in some cases, can deviate from being planar with respect to the ring scaffold.

The findings of this paper should be helpful to better understand the pharmacophore interactions of the flavonoids with the BzR binding site in $GABA_A$ receptors. In future work, compounds having substituents at C16 and C17 may be added to the training set to explore the possible pharmacophore interactions between these parts of the chemical structures and the BzR binding site.

ACKNOWLEDGMENT

Partial funding for this study was provided by National Institutes of Health grant P01-GM 62195. We also gratefully acknowledge support from the Laboratory of Molecular Modeling & Design at UIC and from The Chem21 Group, Incorporated.

REFERENCES AND NOTES

(1) Krogsgaard-Larsen, P.; Frolund, B.; Ebert, B. $GABA_A$ receptor agonists, partial agonists and antagonists. *The GABA receptors*, 2nd ed; Enna, S. J., Bowery, Norman G., Eds.; 1997.

(2) Smith, G. B.; Olsen, R. W. How this takes place is beginning to be understood on a molecular level. The BZD site and the GABA site are both partly located on the α_1 protein subunit of the complex. "Functional domains of $GABA_A$ receptors". *TIPS* **1995**, *16*, 162−168.
(3) Woods, J. H.; Katz, J. L.; Winger, G. Benzodiazepines: use, abuse and consequences. *Pharmacol. Rev.* **1992**, *44*, 151−347.
(4) Medina, J. H.; Viola, H.; Wolfman, C.; Marder, M.; Wasowski, C.; Calvo, D.; Paladini, A. C. Overview − flavonoids: a new family of benzodiazepine receptor ligands. *Neurochem. Res.* **1997**, *22(4)*, 419−425.
(5) Marder, M.; Viola, H.; Wasowski, C.; Waterman, P. G.; Cassels, B. K.; Medina, J. H.; Paladini, A. C. 6-Bromoflavone, a high affinity ligand for the central benzodiazepine receptors is member of a family of active flavonoids. *Biochem. Biophys. Res. Commun.* **1996**, *223(2)*, 384−389.
(6) Haberlein, H.; Tschiersch, K. P.; Schafer, H. L. Flavonoids from *Leptospermum scoparium* with affinity to the benzodiazepine receptor characterized by structure activity relationships and in vivo studies of a plant extract. *Pharmazie* **1994**, *49*, 912−921.
(7) Shen, X.-L.; Nielsen, M.; Witt, M. R.; Sterner, O.; Bergendorff, O.; Khayyal, M. Inhibition of [methyl-3H] diazepam binding to rat brain membranes in vitro by dinatin and skrofulein. *Acta Pharmacol. Sinica*. **1994**, *15*, 385−388.
(8) Dekermendjian, K.; Kahnberg, P.; Witt, M. R.; Sterner, O.; Nielsen, M.; Liljefors, T. Structure−activity relationships and molecular modeling analysis of flavonoids binding to the benzodiazepine site of the rat brain $GABA_A$ receptor complex. *J. Med. Chem.* **1999**, *42*, 4343−4350.
(9) Marder, M.; Estiú, G.; Blanch, L. B.; Viola, H.; Wasowski, C.; Medina, J. H.; Paladini, A. C. Molecular modeling and QSAR analysis of the interaction of flavone derivatives with the benzodiazepine binding site of the $GABA_A$ receptor complex. *Bioorg. Med. Chem.* **2001**, *9*, 323−335.
(10) Huang, X.; Liu, T.; Gu, J.; Luo, X.; Ji, R.; Cao, Y.; Xue H.; Wong, J. T.; Wong, B. L.; Pei, G.; Jiang, H.; Chen, K. 3D-QSAR model of flavonoids binding at benzodiazepine site in $GABA_A$ receptors. *J. Med. Chem.* **2001**, *44*, 1883−1891.
(11) Kahnberg, P.; Lager, E.; Rosenberg, C.; Schougaard, J.; Camet, L.; Sterner, O.; Nielsen, E. Ø.; Nielsen, M.; Liljefors, T. Refinement and evaluation of a pharmacophore model for flavone derivatives binding to the benzodiazepine site of the $GABA_A$ receptor. *J. Med. Chem.* **2002**, *45*, 4188−4201.
(12) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509−10524.
(13) Venkatarangan, P.; Hopfinger, A. J. Prediction of ligand−receptor binding free energy by 4D-QSAR analysis: application to a set of glucose analogue inhibitors of glycogen phosphorylase. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 1141−1150.
(14) Albuquerque, M. G.; Hopfinger, A. J.; Barreiro, E. J.; Alencastro, R. B. Four-dimensional quantitative structure−activity relationship analysis of a series of interphenylene 7-oxabicycloheptane oxazole thromboxane A_2 receptor antagonists. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 925−938.
(15) Ravi, M.; Hopfinger, A. J.; Hormann, R. E.; Dinan, L. 4D-QSAR analysis of a set of ecdysteroids and a comparison to CoMFA modeling. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 1587−1604.
(16) Krasowski, M. D.; Hong, X.; Hopfinger, A. J.; Harrison, N. L. 4D-QSAR analysis of a set of propofol analogues: mapping binding sites for an anesthetic phenol on the $GABA_A$ receptor. *J. Med. Chem.* **2002**, *45*, 3210−3221.
(17) Vedani, A.; Briem, H.; Dobler, M.; Dollinger, K.; McMasters, D. R. Multiple conformation and protonation-state representation in 4D-QSAR: The neurokinin-1 receptor system. *J. Med. Chem.* **2000**, *43*, 4416−4427.
(18) Vedani, A.; McMasters, D. R.; Dobler, M. Multi-conformational ligand representation in 4D-QSAR: Reducing the bias associated with ligand alignment. *Quant. Struct.-Act. Relat.* **2000**, *19*, 149−161.
(19) Streich, D.; Neuburger-Zehnder, M.; Vedani, A. Induced fit − The key for understanding LSD activity. A 4D-QSAR study on the 5-HT2A receptor system. *Quant. Struct.-Act. Relat.* **2000**, *19*, 565−573.
(20) HyperChem Program Release 5.01 for Windows; Hypercube, Inc.; 1996.
(21) Doherty, D. C. MOLSIM User's Guide, The Chem21 Group, Inc., 1780 Wilson Dr., Lake Forest, IL 60045, 1997.
(22) Allinger, N. L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc*. **1977**, *99*, 8127−8134.
(23) Hopfinger, A. J.; Pearlstein, R. A. Molecular mechanics force-field parametrization precedures. *J. Comput. Chem.* **1984**, *5*, 486−492.
(24) Glen, W. G.; Dunn, W. J., III; Scott, D. R. Principal components analysis and partial least squares. *Tetrahedron Comput. Methods*, **1989**, *2*, 349−354.
(25) 4D-QSAR User's Manual, Version 2.0, The Chem21 Group, Inc., 1780 Wilson Dr., Lake Forest, IL 60045, 1997.
(26) Rogers, D. G/SPLINES; A hybrid of Friedman's multivariate adaptive regression splines (MARS) algorithm with Holland's genetic algorithm. The Proceedings of the Fourth International Conference on Genetic Algorithm; San Diego, 1991; pp 38−46.
(27) Rogers, D.; Hopfinger, A. J. Application of genetic function approximation to quantitative structure − activity relationships and quantitative structure − property relationships. *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 854−866.
(28) Friedman, J. Multivariate adaptive regression splines. Technical Report No. 102; Laboratory for Computational Statistics, Department of Statistics, Stanford University, Stanford, CA, November 1988 (revised August 1990).
(29) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening − an overview. *Drug Discovery Today* **1998**, *3*, 160−194.
(30) Hopfinger, A. J.; Reaka, A.; Venkatarangan, P.; Duca, J. S.; Wang, S. Construction of a virtual high throughput screen by 4D-QSAR analysis: application to a combinatorial library of glucose inhibitors of glycogen phosphorylase b. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1151−1160.
(31) Rossi, M.; Rickles, L. F.; Halpin, W. A. The crystal and molecular structure of quercetin: A biologically active and naturally occurring flavonoid. *Bioorg. Chem.* **1986**, *14*, 55−69.
(32) Rossi, M.; Meyer, R.; Constantinou, P.; Caruso, F.; Castelbuono, D.; O'Brien, M.; Narasimhan, V. Molecular structure and activity toward DNA of baicalein, a flavone constituent of the Asian herbal medicine "Sho-saiko-to". *J. Nat. Prod.* **2001**, *64*, 26−31.
(33) Ishiki, H. M.; Alemán, C.; Galembeck, S. E. Conformational preferences of flavone and isoflavone in the gas phase, aqueous solution and organic solution. *Chem. Phys. Lett.* **1998**, *287*, 579−584.

CI0200321