A 3D-QSAR Study of Anticoccidial Triazines Using Molecular Shape Analysis

K.-B. Rhyu,† H. C. Patel, and A. J. Hopfinger*

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

Received March 23, 1995⊗

A set of 54 anticoccidial triazine analogs were analyzed in terms of molecular shape analysis, MSA, with the goal of constructing a three-dimensional quantitative structure−activity relationship (3D-QSAR). This same dataset was previously investigated using alternate QSAR methods including comparative molecular field analysis (CoMFA). A complete 3D-QSAR was realized using MSA which included (a) identification of the active conformation of each individual analog, (b) the relative intramolecular stability of each analog in its active conformation, (c) general intramolecular stability requirements for activity, (d) a statistically significant correlation equation, and (e) mapping of the receptor space explored by the analogs. The MSA-3D-QSAR indicates that the steric shape and dipole moment of each analog governs its activity. The MSA-3D-QSAR correlation equations are more predictive than other QSARs as measured by cross-validation correlation coefficients.

INTRODUCTION

A series of anticoccidial 2-(substituted-phenyl)-1,2,4-triazine-3,5(2*H*,4*H*)-dione, I (the triazines), have been analyzed in terms of quantitative structure−activity relationships, QSARs, using different methods.[1,2] Hammett σ and Hansch π substituent constants were used in the analysis of a small set of simply substituted triazines.[3] The correlation was marginal, the square of the correlation coefficient, R^2, being about 0.5. However, the QSAR strongly suggested lipophilicity was related to potency. This, in turn, led to compounds with potency comparable to that of the commercial drug, diclazuril (**1** in Table 1), which is highly effective in the control of avian coccidiosis.[4]

Subsequently McFarland and co-workers[1] carried out QSAR analyses of the set of anticoccidial triazines given in Table 1 using multiple regression analysis (MRA), the Hansch approach, and linear discriminant analysis (LDA). In that work it was shown the lipophilicities and the acidities of the triazine analogs are important determinants of their anticoccidial potencies. Relative lipophilicity was determined by HPLC, the results being expressed as log k'. The relative acidities were estimated from the ^{1}H-NMR chemical shift at position 6 (δ_6) of the triazine ring. The best QSAR is expressed by eq 1

$$\log (1/\mathrm{MEC}) = -1.20(\pm 0.51)[\log k']^2 + 2.55(\pm 0.52)[\log k'] + 7.49(\pm 2.16)[\delta_6] - 56.57 \quad (1)$$

$$n = 54 \quad R^2 = 0.56, \quad s = 0.71, \quad F_{3,50} = 21.30$$

where MEC is the lowest compound level, in ppms, in feed preventing lesions in a test set of chicks prechallenged with oocysts of the coccidian parasite *Eimeria tenella*.[5] An R^2 value of 0.56 indicates that eq 1 is not sufficiently robust to be predictive, and unrecognized additional contributing factors are involved in the expression of biological activity.

McFarland[2] next applied comparative molecular field analysis (CoMFA) to the data set in Table 1 to determine a three-dimensional QSAR (3D-QSAR). This is the first successful application of CoMFA to result from a whole animal infectious disease model. Still, CoMFA by itself yields no better QSAR models than realized from MRA[1] as expressed by eq 1. However, CoMFA in combination with log k' and δ_6 yields a good 3D-QSAR. The descriptions of these QSAR models are given in Table 2 and taken from ref 2. The "optimum components" column refers to significant partial least-square, PLS, components[6] and the cross-validated R^2 values were computed using the leave-one-out scheme.[7]

A plausible interpretation of these 3D-QSAR models is that physical data, reflecting, for example, transport, may be needed to augment the binding descriptors, the steric and electrostatic field data from CoMFA, to treat *in vivo* biological data. Still, a review of the QSAR analyses of the anticoccidial triazines of Table 1 leads to questions regarding the information extracted from the SAR data. The inclusion of the experimental properties log k' and δ_6 in the 3D-QSAR diminishes the design potential of the QSAR. If an analog must be made to measure some of its physical properties for a QSAR, its biological activity might as well also be measured which negates the predictive impact of the QSAR. Part of the CoMFA study[2] included a correlation between log k' and the log of the water, 1-octanol partition coefficient, log P, as estimated by CLOGP.[8] Substitution of the calculated log k' values from the CLOGP correlation equation could reduce the dependence of the CoMFA 3D-QSAR models on physical property measurements.

An energy minimized structure of diclazuril (**1** in Table 1) was used as the template molecular model in the CoMFA analysis. However, diclazuril and most of the other analogs of Table 1 have considerable conformational flexibility, see Figure 1. Thus, it cannot be resolved if the conformation used in the CoMFA analysis corresponds to the active (receptor-bound) molecular shape. In addition, it is difficult to ascertain the relative intramolecular stability of the structure of each analog used in the CoMFA study without

† On leave from Department of Chemistry, Chongju University, Chongju 360-764, Korea.

⊗ Abstract published in *Advance ACS Abstracts*, June 15, 1995.

0095-2338/95/1635-0771$09.00/0

Table 1. Set of 54 Anticoccidial Triazines Used To Construct the 3D-QSARs and Active Conformations, in Terms of ϕ_1 and ϕ_2, Their Relative Stabilities, ΔE, the x-Component of the Dipole Moment, ΔD_x, and the Observed and Predicted Biological Activities

compd	substituents			ϕ_1	ϕ_2	ΔE	$V_{ov}/10^2$	ΔD_x		log (1/MEC)		Δ[log(1/MEC)]
no.	R_5	R_4	R_3	(deg)	(deg)	(kcal/mol)	(Å^3)	(debye)	log P	obs	pred[a]	residual
1	Cl	$CH(CN)C_6H_4$-4Cl	Cl	60.1	10.3	0.31	2.50	3.59	5.05	3.61	2.91	0.71
2	Cl	SC_6H_4-4-Cl	Cl	59.8	8.2	0.83	2.68	4.25	5.65	3.30	3.76	−0.46
3	Cl	SC_6H_4-4-Cl	Me	50.0	19.9	0.86	2.45	3.45	5.97	3.28	2.66	0.62
4	Cl	SC_6H_4-4-Ac	Me	58.1	8.6	2.29	2.52	1.18	4.77	3.19	2.78	0.41
5	Cl	$CH_2C_6H_4$-4-Cl	Cl	59.8	5.4	0.19	2.42	2.57	6.20	3.18	2.53	0.65
6	Cl	$SO_2C_6H_4$-4-Cl	Cl	54.0	15.0	2.76	2.56	5.86	3.45	2.94	2.57	0.37
7	Cl	COC_6H_4-4-Cl	Cl	58.2	7.1	0.13	2.48	3.96	4.92	2.90	2.75	0.15
8	Me	SC_6H_4-4-Cl	Me	49.7	31.3	0.25	2.45	2.44	6.28	2.86	2.68	0.18
9	Me	$SO_2C_6H_4$-4-Cl	Me	58.7	9.6	1.64	2.53	4.67	4.14	2.59	2.85	−0.26
10	Me	$S(O)C_6H_4$-4-Cl	Me	60.4	8.4	0.98	2.38	3.57	3.80	2.58	2.37	0.21
11	Me	OC_6H_3-3Me-4-SMe	Me	55.6	9.2	0.56	2.42	1.99	6.34	2.27	2.48	−0.21
12	Cl	OC_6H_4-4-Ac	Me	59.9	9.5	1.85	2.39	4.25	4.98	2.27	2.31	−0.04
13	Cl	OC_6H_4-4-CH(OH)Me	Me	62.5	4.6	1.64	2.34	1.86	4.58	2.27	2.12	0.15
14	Cl	SC_6H_4-4-Cl	H	59.6	9.0	0.52	2.48	3.50	5.47	2.26	2.80	−0.54
15	Cl	$CH(OH)C_6H_4$-4-Cl	H	55.7	13.7	0.85	2.37	2.40	3.88	2.26	2.33	−0.07
16	Me	OC_6H_4-4-SMe	Me	56.4	9.2	0.68	2.34	1.16	5.85	2.25	1.97	0.28
17	Me	OC_6H_4-4-SMe	H	55.4	10.4	0.62	2.26	3.24	5.34	2.23	1.94	0.29
18	Cl	OC_6H_3-2-Cl-4-So_2NMeEt	Me	50.0	20.0	1.50	2.40	3.28	5.76	2.11	2.47	−0.36
19	Cl	OC_6H_4-4-Cl	Me	62.9	4.3	1.99	2.26	2.86	6.10	1.98	1.91	0.07
20	Cl	OC_6H_4-4-SMe	H	59.2	10.6	0.87	2.19	1.94	5.42	1.98	1.60	0.38
21	Me	OC_6H_4-4-SO_2Me	H	50.0	19.6	0.54	2.24	4.44	3.42	1.97	1.68	0.29
22	H	$SO_2C_6H_4$-4-Cl	H	62.7	9.4	0.49	2.13	4.65	3.14	1.96	1.29	0.67
23	Cl	H	Cl			0.00	1.68	2.18	3.28	1.81	*b*	*b*
24	Cl	OC_6H_4-4-I	H	60.5	10.7	1.43	2.24	2.95	6.00	1.77	1.85	−0.08
25	Cl	O(naphth-2-yl-6-Br)	H	59.2	9.4	1.36	2.37	2.67	6.88	1.77	2.32	−0.55
26	H	$SO_2C_6H_4$-4-Br	H	60.2	10.5	0.46	2.19	4.60	3.29	1.71	1.49	0.22
27	Cl	OC_6H_3-2,4-Cl_2	H	62.1	15.1	1.39	2.23	1.27	6.06	1.71	1.58	0.13
28	Cl	$SO_2C_6H_4$-4-Cl	Me	54.0	14.6	0.76	2.47	5.71	3.80	1.71	2.20	−0.49
29	Me	OC_6H_4-4-Br	H	59.3	9.4	3.11	2.17	1.86	5.74	1.70	1.53	0.17
30	Me	OC_6H_3-2,4-Cl_2	H	74.8	24.9	2.98	1.87	0.19	5.91	1.69	*b*	*b*
31	Cl	$CH_2C_6H_4$-4-Cl	H	59.3	9.4	0.83	2.31	1.58	5.48	1.64	1.94	−0.30
32	Cl	OC_6H_5-2-Cl-4-SO_2NH-C-C_3H_5	Me	50.0	19.8	2.47	2.41	7.02	5.09	1.51	1.24	0.27
33	Me	$SO_2N(CH_2CH_2)_2O$	Me	72.8	16.2	0.26	2.21	5.07	2.23	1.39	1.43	−0.04
34	Me	COC_6H_4-4-Cl	Me	58.9	9.7	0.57	2.38	1.96	4.99	1.38	2.30	−0.92
35	CF_3	Br	H			0.14	1.71	3.54	3.89	1.35	0.75	0.60
36	CF_3	F	H			0.13	1.54	3.31	3.17	1.26	0.73	0.53
37	Cl	H	Me			0.26	1.66	0.86	3.28	1.20	0.38	0.82
38	Cl	$SO_2N(CH_2OH_2)_2O$	Cl	69.7	19.8	0.32	2.29	5.92	1.54	1.13	1.37	−0.24
39	Cl	$SO_2N(CH_2CH_2)_2O$	H	65.5	11.2	2.28	2.18	5.53	1.39	1.09	1.16	−0.07
40	H	SO_2ph	H	59.5	9.8	0.43	2.11	3.87	2.40	1.04	1.40	−0.36
41	Me	CH(OH)ph	H	59.6	11.2	0.88	2.19	0.53	2.95	1.01	1.21	−0.20
42	CF_3	H	H			0.06	1.51	1.75	3.03	0.93	0.63	0.30
43	Me	H	Me			0.00	1.65	0.58	3.20	0.86	−0.27	1.13
44	CN	H	H			0.19	1.47	1.55	1.56	0.85	0.61	0.24
45	Me	OC_6H_4-4-Cl	H	50.5	8.5	2.52	2.19	1.90	5.59	0.74	1.58	−0.84
46	Cl	S(naphth-2-yl)	H	55.6	7.1	1.07	2.38	2.38	7.81	0.48	*b*	*b*
47	Cl	H	H			0.00	1.53	1.07	2.86	0.25	0.45	−0.20
48	Cl	OMe	Cl	54.8		0.29	1.79	2.25	3.14	0.06	0.80	−0.74
49	H	NO_2	H	0.6		0.06	1.52	5.62	1.85	−0.03	0.21	−0.24
50	H	Oph	H	51.5	6.8	0.18	1.77	0.31	4.21	−0.25	0.25	−0.50
51	H	SO_2Me	H	53.8		0.10	1.72	4.79	0.50	−0.27	0.53	−0.80
52	OMe	H	H			0.00	1.58	0.39	2.21	−0.30	0.19	−0.49
53	H	H	H				1.37	−0.16	2.14	−0.42	0.07	−0.49
54	OMe	OMe	OMe	54.2		0.11	1.73	−0.46	1.53	−0.55	−0.16	−0.39

[a] Values of log(1/MEC) predicted by eq B(1) of Table 3. [b] Outlier.

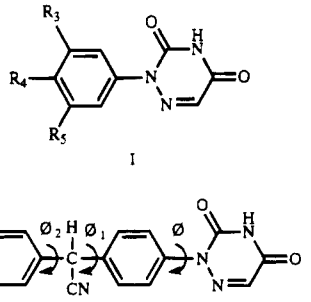
having identified the global minimum energy conformation by complete conformational searching.

Finally, the 3D-QSAR constructed by combining log k' and δ_6 with CoMFA descriptors provides a statistically significant model, but it is not clear how inclusive is the model with regard to the physicochemical descriptors (independent variables) used. One way to probe QSAR uniqueness is to apply alternate methods and descriptor sets as part of the analysis.

The objectives of the study of the anticoccidial triazines reported here are to (1) derive a 3D-QSAR independent of any descriptors which are measured and not computed; (2) identify the active conformations and corresponding molecular shapes of each of the triazine analogs; (3) estimate the relative intramolecular stability of each of the active conformations; and (4) explore 3D-QSARs for the anticoccidial triazines, using molecular shape analysis,[9] MSA, and identify the most significant models.

METHODS

1. Biological Activity. The same set of anticoccidial activities, expressed as log (1/MEC) measures, used in the QSAR studies reported in refs 1 and 2 were used in this work and are given as part of Table 1.



I

Figure 1. The structure of diclazuril with the torsion angles (ϕ, ϕ_1, ϕ_2) defined.

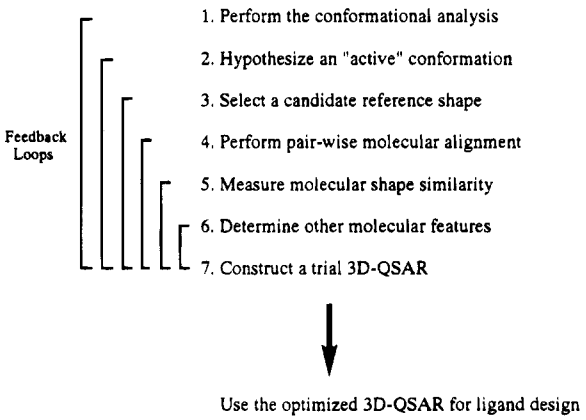


Figure 2. The basic operations of molecular shape analysis.

2. Building Analogs. The general form of the anticoccidial triazines considered in this study are given by I and the major torsion angles ϕ, ϕ_1, and ϕ_2 are given for diclazuril in Figure 1. The bond lengths and angles of the triazine ring were taken from the crystal structure of 6-azauracil.[10] All other bond lengths, bond angles, and torsion angles were fully optimized for each analog using the MNDO Hamiltonian of MOPAC.[11] MNDO was not used to compute the trial QSAR molecular descriptors described below. The triazine ring geometry was frozen in each optimization, because the crystal structure was judged to be adequate when comparing it to the MNDO optimized structure using I. Further, the triazine ring is common to all analogs and is thus a "constant" to all QSAR analyses of the SAR data base.

3. Molecular Shape Analysis, MSA. The current formulation of MSA involves seven interactive operations which are listed in Figure 2 and summarized for this study as follows.

A. Conformational Analysis. The SCAN option of CHEMLAB-II[12] was used to perform a fixed valence geometry search of the principal torsion angles ϕ, ϕ_1, and ϕ_2 at 10° increments. The torsion angles of analogs with flexible substituents, in addition to ϕ, ϕ_1, and ϕ_2, were explored at successive refinement searches of 60°, 30°, and 10° resolution. The reference conformation $\phi = \phi_1 = \phi_2 = 0°$ corresponds to the planes of the three rings of diclazuril being coplanar.

A fixed valence geometry molecular mechanics force field composed of dispersion/steric, electrostatic, hydrogen-bonding, and torsion potential contributions was used to estimate the conformational energy. The nonbonded potential set proposed by Hopfinger[13] was used to compute the dispersion/steric interactions. The electrostatic interactions were calculated using a Coulomb potential with a molecular dielectric of 3.5 and MNDO-calculated atomic charges. The hydrogen-bonding potential developed by Hopfinger[13] was used. Torsion potential functions were selected, or derived, based on the parameterization procedures given in refs 12 and 14.

All apparent intramolecular energy minima, that is minima defined by the resolution ($\Delta\phi_1$, $\Delta\phi_2$) of the conformational search, within 6 kcal/mol of the apparent global minimum, were identified and recorded for each compound. Each of the apparent minima were then used as starting points in rigorous fixed valence geometry energy minimizations (zero convergence of the energy-torsion angle first derivatives). Both the apparent and the rigorous sets of minimum energy conformations were used as candidates for the "active" conformation of each analog in subsequent MSA steps.

The conformation assigned to a flexible substituent was selected to satisfy the maximum destabilization energy constraint, and, at the same time, maximize molecular shape similarity to the molecular reference shape. Conformational stability and shape similarity are discussed below.

B. Selection of the Active Conformation and Reference Shape. The LBA−LCS (loss in biological activity−loss in conformational stability) strategy was employed to identify the active conformation of each analog. This approach attempts to identify stable intramolecular conformer states for active analogues that are not stable for inactive analogs. The premise implicitly made is that the loss in activity for the inactive analogs is a result of not being able to adopt the "active" conformation energetically available to the active compounds. Different upper energy cutoffs, relative to the global minimum, were used to define the sets of candidate active conformations from the sets of intramolecular minima. The statistical significance of the 3D-QSAR was used to select the preferred energy cutoff. The optimal cutoff energy was found to be 3 kcal/mol above the apparent global minimum.

The search for the active conformation began with a small number of compounds from Table 1, namely 2, 3, 19, and 45. These compounds were selected to include very active, moderately active, and inactive analogs. The active conformation was postulated by observing which apparent conformational energy minimum common to the active analogs, compound 2, 3, and 19, was not available to the inactive analog, compound 45. The postulated active conformation was then assigned to successively larger subsets of the compounds in Table 1 until the LBA−LCS strategy was violated or until all analogs of Table 1 were successfully modeled. The initial candidate for the active conformation was the apparent global minimum of diclazuril and proceeded to increasingly higher energy, but stable, intramolecular conformer candidates.

Once a common apparent local minimum energy conformer was identified as a likely active conformation for a set of analogs, the corresponding set of rigorous minimum energy conformers corresponding to this minimum were considered as the individual active conformations of each respective analog. Only one apparent local minimum was found to satisfy the LBA−LCS strategy for all analogs of Table 1. The intramolecular energy relative to the global minimum, ΔE, of each analog, in its active conformer state, was retained as an independent descriptor in the 3D-QSAR analysis.

All conformations of all analogs in Table 1 and, more generally, analogs outside the training set (Table 1) are viable candidate reference shapes. However, it makes sense that

the more active analogs in the training set possess necessary, but not necessarily all, shape requirements for activity. Thus, this reasoning can be invoked as a constraint in the search for an optimal reference shape. In this study the reference shape search was restricted to analogs of Table 1, in their postulated active conformations, which have at least 80% of the log (1/MEC) activity of analog 1 (log (1/MEC) ≥ 2.90 which includes compounds **1**−**7** in Table 1). This is an arbitrary, but plausible, design constraint. Compound **2** as the reference shape was found to yield the statistically most significant 3D-QSAR.

C. Molecular Alignment. MSA requires that each compound in the data set is compared to the reference shape. Such a comparison necessitates a pairwise molecular alignment. The geometric criterion for pairwise analog molecular alignment was to place the N−N−C atoms of the triazine ring of each pair of molecules identically upon one another. The reason that these atoms were chosen was because they insure that the largest common substructure of the analogs is given a common consideration in the shape comparisons.

D. Quantitative Measures of Molecular Shape. Two descriptors which measure relative molecular shape similarity were considered. The common overlap steric volume, V_{ov},[15] between each analog, u, in the data set and the shape reference compound, v, was determined as

$$V_{ov} = V_u \cap V_v \tag{2}$$

where V_u and V_v are the volumes of the isolated molecules, respectively. V_{ov} was computed using a numerical integration scheme in the QSAR module of the Drug Discovery Workbench[16] and is a measure of how similar in steric shape the analogs are to the shape reference.

The other molecular shape similarity descriptor considered in this investigation is the nonoverlap volume, V_{non},[17] defined as

$$V_{non} = V_{uv} - V_v \tag{3}$$

where V_{uv} is the composite steric volume of the aligned u,v pair. The nonoverlap volume measures the regions of space of u not shared by the two molecules. V_{ov} and V_{non} do not contain information regarding the intramolecular stability of each conformation of the pair of compounds from which they are derived.

E. Selection and Estimation of Other Molecular Descriptors. The molecular decomposition-recombination, MDR, technique[18] was used to determine which nonshape physicochemical properties might be potential molecular descriptors in the MSA-3D-QSAR. The MDR technique assumes that a molecule can be decomposed into a set of "substructure molecules" such that each set of individual substructure molecules can be analyzed, in terms of QSARs, and/or molecular modeling, using the structure−activity data of the corresponding parent (whole) compounds. Clearly, the MDR assumption is predicated upon selecting specific parent molecular structure−activity data in which the only changes in chemical structure occur in the substructure of interest. The remainder of the parent molecule remains constant. Further, each substructure of the whole molecule does not modify the properties of any other substructure, that is, the substructures are uncoupled from one another with respect to computation of the molecular descriptors of interest.

In this study, the QSAR analysis is based upon a molecular substructure that is bonded to the remainder of the parent compound (the triazine ring) by a saturated single C−N bond. Thus, minimal resonance/electronic substructure complications should occur in the estimation of electronic descriptors, characteristic of the parent compound, by using its substructure representative in a MSA-3D-QSAR analysis.

i. Relative Lipophilicity (LOGP). The lipophilicities, LOGP, of the non-triazine substructures were determined using the software package CLOGP.[8] These LOGP values should be proportional, over the data set, to the lipophilicities of the whole molecules since the triazine ring is common to all analogs.

ii. Dipole Descriptors (ΔD_x, ΔD_y, ΔD_z, ΔD_t). The total dipole and the x, y, and z components (based upon the bond from the triazine ring to the R_4 substituent aligned with the positive x-axis, and the triazine ring in the xy plane) were computed for each analog in its respective active conformation using the CNDO/2 approximation.[19] The MDR technique was applied to determine the non-triazine ring dipole contributions ΔD_x, ΔD_y, ΔD_z, and ΔD_t from their respective whole molecule x, y, z, and total, t, components.

iii. Electronic Descriptors (HOMO, LUMO). The highest occupied and lowest unoccupied molecular orbital energies (HOMO and LUMO), calculated by CNDO/2 method, were considered as molecular descriptors.

iv. Relative Conformational Stability Descriptor (ΔE). The relative conformational stability, ΔE, defined as the difference between a particular conformation and the global intramolecular minimum energy conformation was considered as a descriptor.

F. Construction of Trial MSA-QSAR. Trial MSA-3D-QSARs were constructed using two statistical analysis strategies and corresponding methods.

i. Combinatorial, with respect to the molecular descriptors, multidimensional linear regression analysis was used to generate trial QSAR models. Both linear and quadratic descriptor representations were considered with the constraint that no pair of descriptors in a QSAR model could have a cross-correlation coefficient greater than 0.50. Analogs were considered as outliers and removed from the regression analysis when the difference in predicted and observed activities exceeded 2.0 standard deviations from the mean. The partial F-test was applied to each regression equation to determine whether the addition of any particular descriptor term to a QSAR model significantly contributed to the prediction of the activity. The facilities of the SAS software package[20] were used to construct trial QSAR models.

ii. A genetic algorithm, GA, scheme, called the GFA,[21] genetic function approximation, available in the QSAR module of the Drug Discovery Workbench, which includes power splining,[21] was used to evolve and optimize MSA-3D-QSAR models. The procedure used to evolve the QSAR models for the Cardozo−Hopfinger database in ref 21 was applied in this study.

The predictive robustness of all QSAR models was evaluated using leave-one-out cross-validation[7] available in the Drug Discovery Workbench.

RESULTS

The MSA-3D-QSAR models are virtually independent of conformation as expressed by torsion angle ϕ. Further, the

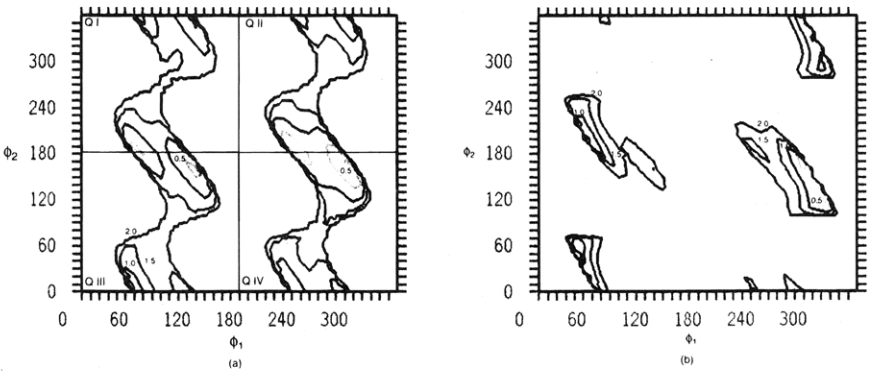


Figure 3. The (ϕ_1, ϕ_2) conformational energy maps: (a) compound **2** and (b) compound **3**. Energy contours are in kcal/mol relative to the global minimum and the upper energy contour cutoff is 2 kcal/mol.

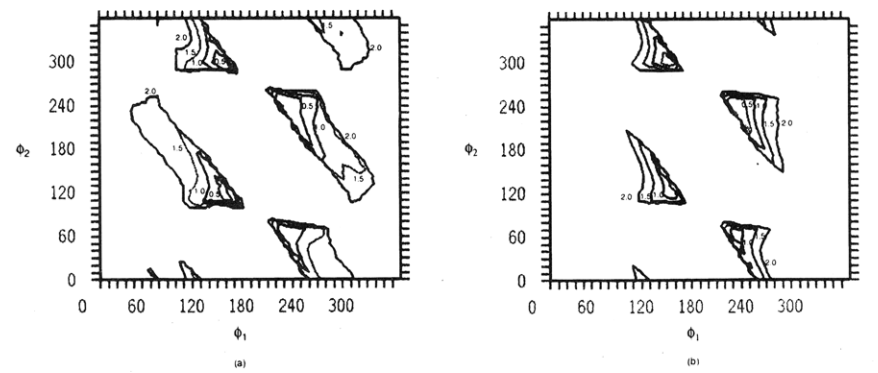


Figure 4. The (ϕ_1, ϕ_2) conformational energy maps of (a) compound **19** and (b) compound **45**.

torsion angle ϕ energy profile is extremely similar for all analogs in Table 1. Consequently, the global intramolecular minimum energy conformation of compound **1**, with respect to ϕ, was assigned to all analogs.

The (ϕ_1, ϕ_2) conformational energy maps for analogs 2 and 3 are shown in Figure 3. These two compounds are both of about equally high activity. Thus, these two analogs should be able to adopt the active conformation with similar relative stability. Different maximum ΔE values were considered in defining the set of intramolecular energy minima as candidates for the active conformation. The most significant MSA-3D-QSARs are realized for maximum ΔE values in the 2−5 kcal/mol range. The optimum maximum ΔE value in the analysis is 3 kcal/mol. Overall, the set of relative minima for analog 2 define the set of active conformation candidates in terms of ϕ_1 and ϕ_2. The (ϕ_1, ϕ_2) energy maps for the moderately active compound 19, and the inactive analog 45, are shown in Figure 4.

No significant MSA-3D-QSARs could be constructed using any of the *apparent* intramolecular relative minima, as defined by compound **2** and the 2−5 kcal/mol maximum ΔE cutoff. However, when the corresponding set of *rigorous* torsion angle minima are used with the 2−5 kcal/mol maximum ΔE cutoff, some significant MSA-3D-QSARs can be found. In particular, rigorous minima derived from the (ϕ_1, ϕ_2) = (60°, 10°) apparent relative minimum provide a base for very significant MSA-3D-QSAR models.

Compounds **2**, **3**, and **7** each lead to significant MSA-3D-QSARs when used as the reference molecular shapes. Compound **2**, in its postulated active conformation, yields the optimum QSAR model when used as the reference shape.

The optimum MSA-3D-QSARs, using the (ϕ_1, ϕ_2) = (60°, 10°) conformation and compound **2** as the reference molecular shape, are given in Table 3. The specific (ϕ_1, ϕ_2) and

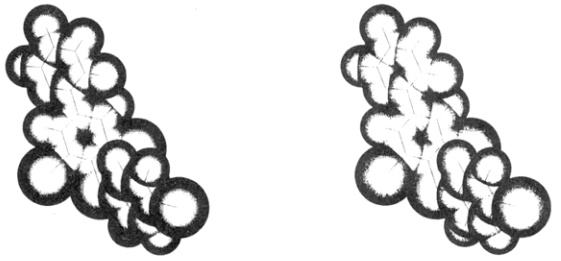


Figure 5. A stereo space-filling and stick representation of compound **2**, the reference shape, in its postulated active conformation (ϕ_1, ϕ_2) = (59.8°, 8.2°).

Table 2. Comparison of Hansch-Type and CoMFA QSAR Models, Taken from Ref 3

descriptors	optimum PLS components	N	R^2	xv-R^2	F	SD
A. $(\log k')^2 \log k' \delta_6$	3	54	0.561	0.488	21.3	0.71
B. CoMFA field only	2	54	0.658	0.469	*a*	*a*
C. A and B combined	4	54	0.801	0.613	49.2	0.48

[a] Not reported.

ΔE values for the individual active conformations of each analog are given as part of Table 1. The reference molecular shape (the active conformation of compound **2**) is shown in Figure 5.

The three MSA-3D-QSARs of Table 3 are for the same model, in one case, (A), including all 54 compounds, and delete the three outliers in the other two cases, B(1) and B(2). This model was found in both the combinatorial MRA and in GA statistical analyses. The squared correlation and cross-validation correlation coefficients are R^2 and xv-R^2, respectively. The number of the compounds, N, outliers by compound number, maximum cross-correlation coefficient among the descriptors, xC_{max}, standard deviation of fit, SD,

Table 3. Optimum MSA-3D-QSAR Equations Found in the Study[a]

MSA-3D-QSARs	V_{ov}(ext) (Å^3)	ΔD_x(ext) (debye)	R^2	xv-R^2	F	SD	xC_{max}	outlier no.	N
(A) $\log(1/\text{MEC}) = -7.03(\pm3.32)V_{ov} + 2.26(\pm0.82)V_{ov}^2 + 0.42(\pm0.15)\Delta D_x - 0.07(\pm0.02)\Delta D_x^2 + 5.63(\pm3.24)$	155.5	3.00	0.70	0.64	28.7	0.59	0.42		54
(B(1)) $\log(1/\text{MEC}) = -8.42(\pm2.77)V_{ov} + 2.65(\pm0.69)V_{ov}^2 + 0.48(\pm0.13)\Delta D_x - 0.08(\pm0.02)\Delta D_x^2 + 6.71(\pm2.64)$	158.9	3.00	0.81	0.79	47.8	0.48	0.38	23, 30, 46	51
(B(2)) $\log(1/\text{MEC}) = -8.05(\pm2.65)V_{ov} + 2.48(\pm0.66)V_{ov}^2 + 0.46(\pm0.13)\Delta D_x - 0.07(\pm0.02)\Delta D_x^2 + 5.58(\pm2.51)$	162.3	3.28	0.83	0.81	36.0	0.46	0.63	23, 30, 46	51

[a] Equation A is for all 54 analogs, while eqs B(1) and B(2) are derived with the three outliers of eq A deleted from the analog training set. Outlier no. refers to the compound numbers given in Table 1. The log P and log P^2 terms in eq B(2) are not significant by the partial F test criterium but are included for comparison to the QSAR models of Table 2.

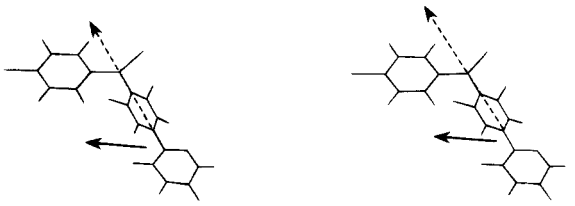


Figure 6. A stereo stick representation of diclazuril (compound **1**) in its active conformation with ΔD_t and ΔD_x superimposed.

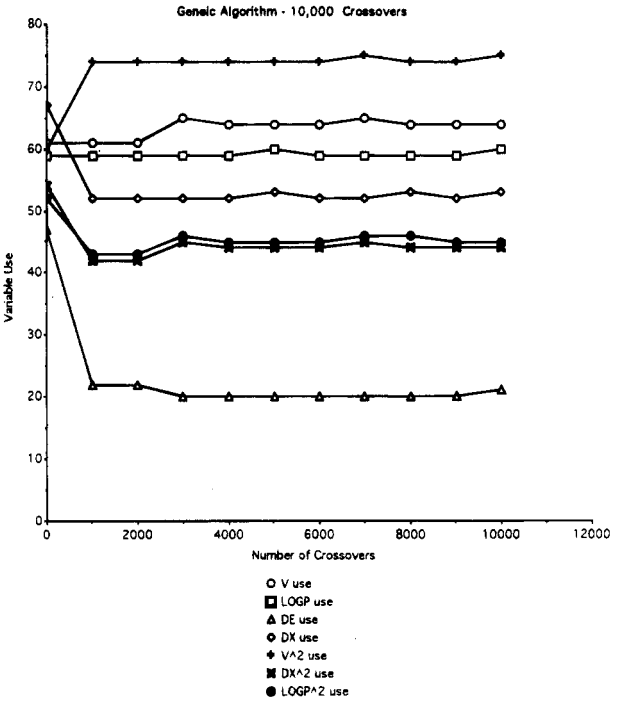


Figure 7. A plot of number of crossovers in the GA versus the number of times a variable (descriptor) has been used to that point in the evolution. GA health is measured by the LOF measure[21] and not R^2. This is for all 54 compounds.

and the F statistic of significance are also given in Table 3. The two QSAR descriptors, V_{ov} and ΔD_x, of the MSA-3D-QSARs in Table 3 both occur in linear and square terms. The corresponding extremum values for these two descriptors are reported in Table 3. Equation B(2) of Table 3 includes log P and log P^2 terms. The log P descriptor terms are not statistically significant by the partial F-value test. However, they are included to permit comparisons to the QSAR models reported in Table 2.

The parabolic relationship between log(1/MEC) and ΔD_x is plausible both from a statistical and a physicochemical point of view. The MSA-3D-QSAR B(1) of Table 3 suggests log(1/MEC) is optimized for a ΔD_x of about 3.00 debye. This could be a reflection of a specific electrostatic matching interaction between ligand and receptor. Twenty-five analogs in the data set have larger ΔD_x values than ΔD_x-(opt), and 29 analogs have smaller ΔD_x values. This indicates the parabolic behavior of ΔD_x is distributed over the entire set of analogs.

In contrast to ΔD_x, the parabolic dependence between log-(1/MEC) and V_{ov} appears to be a data-fitting requirement. The MSA-3D-QSAR B(1) of Table 3 indicates that log(1/MEC) is *minimized* for a V_{ov} of about 158.9 Å^3. Only six relatively inactive analogs have V_{ov} < 158.9 Å^3, and each of these analogs has a V_{ov} measure close to this extremum value (137.0 to 158.9). The other 48 analogs cover a much

Table 4. Three Outliers Found in the 3D-MSA-QSAR Analysis

compd no.	R_5	R_4	R_3	log(1/MEC) obs	log(1/MEC) pred[a]	log(1/MEC) diff	V_{ov} (Å^3)	ΔD_x (debye)
23	Cl	H	Cl	1.81	0.78	1.03	167.7	2.18
30	Me	OC_6H_3-2,4-Cl_2	H	1.69	0.47	1.22	186.9	0.19
46	Cl	S(napthyl-2-yl)	H	0.48	2.29	−1.81	238.0	2.38

[a] Based on eq A of Table 3.

larger range of V_{ov} values (158.9 to 268.0). Thus, the parabolic behavior of V_{ov} in the QSARs largely accounts for a somewhat nonlinear relationship between V_{ov} and log(1/MEC) for the moderately and highly active analogs.

DISCUSSION

The combined physical measurement and CoMFA 3D-QSAR, model C of Table 2, has a more significant R^2 value, 0.80, than does the equivalent 3D-MSA-QSAR, model A of Table 3, $R^2 = 0.70$. However, the xv-R^2 of the MSA model, 0.64, is modestly better than that of the CoMFA model, 0.61. There are some other attractive aspects of the MSA-based QSARs,

1. An active conformation is determined for each analog. The active conformation, in terms of ϕ_1 and ϕ_2, varies modestly from analog to analog. The conformational variation is done to minimize conformational energy, yet retains V_{ov} and ΔD_x values which yield a significant correlation equation. It is not possible to quantitatively compare the MSA active conformations to the conformation used in the CoMFA 3D-QSAR study. However, a visual comparison of the conformation used for compound **1** (diclazuril), as reported in the CoMFA study,[2] to the postulated MSA active conformation, shown in Figure 6, suggests the triazine ring and the R_4 = $CH(CN)C_6H_4$-Cl substituent are both rotated out of the plane of the central phenyl ring to a greater extent, in the CoMFA conformation, as compared to the MSA conformation. The conformational implications to field potential cannot be ascertained.

2. Requirements for intramolecular ligand stability for the active conformation have been identified.

3. Only two molecular descriptors, both of which can be easily computed, are needed to construct the 3D-QSAR correlation equations.

4. The consideration of many possible molecular descriptors and power splines in the GA analysis gives confidence that the best models found in the study are, indeed, significant.

In contrast to the Hansch-type and CoMFA QSARs, lipophilicity, as measured by log P in this study, is not found to be a significant QSAR descriptor in the MLR analyses. The inclusion of log P, and (log P)^2, which is found to be the best functional representation of this descriptor if it is used, only increases the R^2 of QSAR B(1) of Table 3 from 0.81 to 0.83, see QSAR B(2) in Table 3. The correlation contributions of lipophilicity seen in the Hansch-type and CoMFA QSARs appear to be inherent to V_{ov}. There is a relatively large cross-correlation coefficient between V_{ov} and log P of 0.64. The GA studies also suggest that V_{ov} and log P share much of the same QSAR information. Figure 7 is a plot of the evolution of a GA study, expressed as the number of crossovers versus how often a particular descriptor representation has been found over the previous generations (variable use). It seems clear that V_{ov}^2 and V_{ov} are of dominant use in model evolution but that log P survives as the third most often used descriptor. Log P appears to be used preferentially when V_{ov}^2 and/or V_{ov} is not employed in a model. The plateau behavior of all of the descriptor representations shown in Figure 7 suggests a type of equilibrium is realized by 4000 crossovers.

Three outliers are found in the best MSA-3D-QSAR model which are reported in Table 4. If these outliers are not included in the analysis, then the highly significant MSA-3D-QSAR correlation equation given by B(1) in Table 3 is realized. The value, 0.79, for xv-R^2 indicates this is a very reliable QSAR equation in terms of predictivity.

No definitive reasons can be given as to why the three outliers, see Table 4, do not fit the MSA-3D-QSAR model expressed by eq A of Table 3. However, it is noted that compound 46 has R_4 = S (napthyl-2-yl) which is a larger planar ring system than the R_4 of any other analog. This substituent spans a region of space not explored by the other compounds. Compounds **23** and **30** are both about ten times more active than predicted by eq A. These two compounds do not share any common substituent at a given site, but their substituents are common to other well-predicted analogs in the training set. Clearly, the descriptors V_{ov} and/or ΔD_x, or their mathematical representation, is/are not adequate to account for the high activity of these two analogs.

The three outliers are relatively inactive analogs which indicates that the MSA-3D-QSAR models given in Table 3 fit/explain the SAR of all active analogs. Interestingly, GA splining does not suggest QSAR models in which the complete SAR of Table 1 is divided into two subsets (active and inactive analogs) and individual models developed for each subset. Overall, the QSAR models of Table 3 are significant in spanning the entire training set of 54 analogs, but do a somewhat better job fitting the high activity SAR than the less active SAR behavior.

In summary, the MSA-3D-QSAR models reported in this paper offer an alternative interpretation of structure−activity data in Table 1 as compared to previous QSARs. The MSA models also provide additional information on the active conformations of the analogs not realized by other studies. The MSA correlation equations have fewer descriptors than other proposed equations, all descriptors can be computed, and the statistical reliability of the MSA-3D-QSAR model, eqs A and B(1), is at least moderately superior to any other proposed QSAR.

ACKNOWLEDGMENT

This work was performed using the resources of the Laboratory of Molecular Modeling and Design at UIC. K.-B. Rhyu acknowledges the sabbatical fellowship support of Chongju University. We also appreciate the help and useful discussions with J. Tokarski, M. Ravi, and D. Singh of our Laboratory. D. Rogers, D. Swindell, and P. Gund of

Molecular Simulations Inc. made valuable suggestions to us over the course of this study.

REFERENCES AND NOTES

(1) McFarland, J. W.; Cooper, C. B.; Newcomb, D. M. *J. Med. Chem.* **1991**, *34*, 1908.
(2) McFarland, J. W. *J. Med. Chem.* **1992**, *35*, 2543.
(3) Miller, M. W.; Mylari, B. L.; Howes, H. L., Jr.; Figdor, S. K.; Linch, M. J.; Linch, J. E.; Gupta, S. K.; Chappel, L. R.; Koch, R. C. *J. Med. Chem.* **1981**, *24*, 1337.
(4) Vanparijs, O.; Marsboom, R.; Desplenter, L. *Poultry Sci.* **1989**, *68*, 489.
(5) Chappel, L. R.; Howes, H. L.; Linch, J. E. *J. Parasitol* **1974**, *60*, 415.
(6) Dunn, W. J., III; Wold, S.; Edlund, U.; Hellberg, S.; Gasteiger, J. *Quant. Struct.−Act. Relat.* **1984**, *3*, 131.
(7) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E.; Frank, I. E. *Quant. Struct.−Act. Relat.* **1988**, *7*, 18.
(8) Medicinal Chemistry Software, *Medchem Software Manual, Release 3.51*; Pomona College: Claremont, CA, April 1987.
(9) Hopfinger, A. J.; Burke, B. J. *3-D QSAR in Drug Design-Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM Science: Leiden, Netherlands, 1993; p 276.
(10) Singh, P.; Hodgson, D. J. *Acta Crystallogr.* **1974**, *B30*, 1430.
(11) Stewart, J. J. P.; Seiler, F. K. QCPE no. 455, V 4.0; Indiana University, Bloomington, IN 47405, 1987.
(12) CHEMLAB-II VII.0; Molecular Simulations Inc.: 16 New England Executive Park, Burlington, MA, 01803, 1992.
(13) Hopfinger, A. J. *Conformational Properties of Macromolecules*; Academic Press: New York, 1973.
(14) Hopfinger, A. J.; Pearlstein, R. A. *J. Comput. Chem.* **1984**, *5*(5), 486.
(15) Burke, B. J.; Hopfinger, A. J. *J. Med. Chem.* **1990**, *33*, 274.
(16) The Drug Discovery Workbench; Molecular Simulations Inc.: 16 New England Executive Park, Burlington, MA 01803, 1995.
(17) Tokarski, J. S.; Hopfinger, A. J. *J. Med. Chem.* **1994**, *37*, 3639.
(18) Hopfinger, A. J. *J. Am. Chem. Soc.* **1980**, *102*, 7196.
(19) Pople, J. A.; Beveridge, D. C. *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970.
(20) *SAS User's Guide: Basics, SAS Release 5.18*; SAS Institute, Inc.: Cary, NC, 1986.
(21) Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.

CI950027F