Polarizability Fields for Use in Three-Dimensional Quantitative Structure−Activity Relationship (3D-QSAR)

Mary Bradley*

Scynexis Chemistry and Automation, Inc., P.O. Box 12878, Research Triangle Park, North Carolina 27709

Chris L. Waller

Sphinx Pharmaceuticals, A Division of Eli Lilly and Company, 20 T.W. Alexander Drive, Research Triangle Park, North Carolina 27709

Comparative molecular field analysis (CoMFA), a three-dimensional quantitative structure−activity relationship (3D-QSAR) technique, has proven to be a valuable tool in the field of rational drug design. In its native form, CoMFA utilizes a pseudoreceptor, in the form of a regularly spaced lattice of probe atoms, to characterize the steric and electrostatic properties of a series of mutually superimposed molecules. Statistical analyses are performed in an attempt to correlate changes in these shape and charge related fields to observed differences in biological activities at a given target. Graphical analyses of the resulting "negative receptor images" have been demonstrated to provide insight into the physicochemical requirements of novel ligands. Several groups have previously demonstrated the benefits of additional or alternative fields for these types of analyses. In this report, a novel molecular potential field derived from atomistic contributions to molecular polarizability is presented. Comparison studies will be presented using literature data sets and CoMFA models derived from steric, electrostatic, and polarizability fields. The overall conclusion is that molecular polarizability fields derived from semiempirically determined atomic polarizabilities are highly predictive and graphically descriptive supplements to, and perhaps surrogates for, the standard CoMFA steric and electrostatic fields.

INTRODUCTION

Comparative molecular field analysis (CoMFA)[1,2] is a three-dimensional quantitative structure−activity relationship (3D-QSAR) technique in which molecules are represented by their steric and electrostatic fields computed on a pseudoreceptor of regularly spaced probe atoms. The series of molecules under consideration must be superimposed prior to the analysis using any of a variety of manual or automated molecular alignment techniques, ranging from simple atom-pair based root-mean-squares (RMS) fitting to techniques that optimize the mutual alignment based on molecular field similarites (e.g. SEAL[3]). The differences in these field values (as potential energies) are then correlated to observed differences in a target property (e.g. receptor binding affinity) and statistically crossvalidated so as to indicate the likelihood that the results have predictive validity. The properties of the resulting regression equations may then be mapped onto the fields and represented graphically for analysis.

Traditionally, these steric and electrostatic fields have been used exclusively to represent the properties of molecules; however, additional fields including hydrophobicity (HINT),[4] hydrogen bonding (GRID),[5] and molecular orbital (MO-PAC)[6,7] have been developed and implemented with success. These fields expand upon the original principles of steric and electrostatic interactions to define molecular interactions and can result in models that are more predictive for certain applications. We have developed a new field type of field that describes the polarizable character of molecules and can be used alone or in conjunction with other field types in CoMFA studies.

It is theorized that polarizability may play an important role in understanding the interaction of molecules at a distance, especially when permanent dipoles are weak and ionization energies are consistent. As an example, it has been observed that the degree and position of chlorination of polychlorinated biphenyls (PCB) and dibenzodioxins (PCDD) affect the binding of these molecules to the arylhydrocarbon (Ah), or dioxin, receptor and subsequent biological activities (e.g. induction of arylhydrocarbon hydroxylase activity).[8] McKinney et al. proposed a model to account for these differences based on polarizabilities[9] calculated according to a modified version of the Applequist-Carl-Fung (ACF) method based on atom dipole interactions.[10] In this report, it was noted that for PCBs, the long axis component of polarizability was insensitive to rotation about the twist bond (in the plane of the long axis), while the remaining two components varied greatly with changes in the torsion of the twist bond. However, throughout the entire range of available torsion angles, the total molecular polarizability varied minimally (estimated at only 2%).

McKinney et al.[8] provided theoretical evidence to support the hypothesis that for similarly sized and shaped molecules with equal access to the same receptor molecule, polarizability through dispersion would be the determining factor

* Corresponding author phone: (919)544-8608; e-mail: mary.bradley@scynexis.com.

for binding. It follows that for molecules with similar polarizabilities and different sizes and shapes (e.g. non-*ortho*-substituted and *ortho*-substituted PCBs) that other factors must be considered. For the two PCB examples above, the approach distance to the receptor would be greater for the *ortho*-substituted molecule due to the noncoplanar orientation of the two phenyl rings. It is reasonable to assume that an *ortho*-substituted PCB would display weaker affinity for the receptor than a similar non-*ortho*-substituted congener. An increase in the total polarizability of the *ortho*-substituted PCB relative to the non-*ortho*-substituted PCB would be required to achieve a similar degree of binding affinity.

To explore to the spatial and electrostatic requirements of ligands for the Ah receptor, Waller and McKinney used CoMFA to obtain a 3D-QSAR model for a set of poly-halogenated dibenzo-p-dioxins, dibenzofurans, and biphenyls.[11] In this analysis, standard CoMFA steric and electrostatic fields were utilized and were determined to be sufficient to describe the receptor binding affinities of these compounds. Additionally, the model was able to account, in part, for the importance of the degree and substitution pattern of chlorination that had been observed by Safe.[8] The receptor approach distance hypothesis suggested earlier by McKinney was supported by this work in that all *ortho*-substituted biphenyls were treated as noncoplanar molecules in the best 3D-QSAR models that were derived. Subsequently, Waller and McKinney[12] expanded this model to include two additional sets of molecules (i.e. indolocarbazoles and naphthalenes). In this latter study, it was determined that model interpretability was enhanced by the addition of hydropathic interaction, HINT, fields.

Given the relative successes of these two somewhat disparate methods (i.e. polarizability and CoMFA) for creating predictive and interpretive models for diverse ligands binding to the Ah receptor, herein a technique is proposed that combines both principles. Since molecular polarizability has the property of being additive and constitutive based on atomistic values, the straightforward nature of the CoMFA technique makes it amenable to calculation and inclusion as an additional molecular field. As a practical example, 3D-QSAR models were developed and presented herein that not only are statistically robust and predictive but also give more accurate insight into the binding of aromatic compounds to the Ah receptor.

METHODS

Materials and Approaches. The molecular modeling and CoMFA studies were carried out on a Silicon Graphics Octane workstation running Sybyl 6.5[13] CoMFA was used as the three-dimensional QSAR method to analyze the correlation between the calculated fields and the biological activities. Partial least squares (PLS) regression[14] was used in conjunction with leave-one-out (LOO) crossvalidation in order to measure the internal consistency and, nominally, the predictive ability of the resulting QSAR models.

Data Set. The data set consists of 25 polychlorinated and polybrominated dibenzo-p-dioxins, 14 polychlorinated biphenyls, 39 polychlorinated dibenzofurans, 5 polybrominated naphthalenes, and 16 indolocarbazoles and derivatives thereof. All 99 compounds were assayed for their ability to bind to the cytosolic Ah receptor in competition binding

Table 1. Dibenzo-p-dioxins Included in the CoMFA/QSAR Analyses
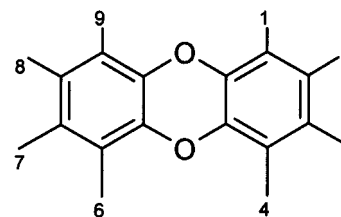
no.	R1	R2	R3	R4	R6	R7	R8	R9	pEC$_{50}$
1	H	Cl	Cl	H	H	Cl	Cl	H	8.000
2	Cl	Cl	Cl	H	H	Cl	Cl	H	7.102
3	H	Cl	Cl	H	Cl	Cl	H	H	6.796
4	H	Cl	Cl	H	Cl	H	H	H	6.658
5	Cl	Cl	Cl	Cl	H	Cl	Cl	H	6.553
6	Cl	H	Cl	H	H	Cl	Cl	H	6.102
7	Cl	Cl	H	Cl	H	Cl	Cl	H	5.959
8	Cl	Cl	Cl	Cl	H	H	H	H	5.886
9	H	Cl	Cl	H	H	Cl	H	H	7.149
10	H	Cl	H	H	H	H	Cl	H	5.495
11	Cl	Cl	Cl	Cl	H	Cl	H	H	5.194
12	Cl	Cl	H	Cl	H	H	H	H	4.886
13	Cl	Cl	Cl	Cl	Cl	Cl	Cl	Cl	5.000
14	Cl	H	H	H	H	H	H	H	4.000
15	H	Br	Br	H	H	Br	Br	H	8.824
16	H	Br	Br	H	H	Cl	Cl	H	8.830
17	H	Br	Cl	H	H	Cl	Br	H	9.350
18	H	Br	Cl	H	H	Cl	Cl	H	7.939
19	Br	H	Br	H	H	Br	H	Br	7.032
20	Br	H	Br	H	H	Br	Br	H	8.699
21	Br	Br	H	Br	H	Br	Br	H	7.770
22	Br	Br	Br	H	H	Br	Br	H	8.180
23	H	Br	Br	H	H	Br	H	H	8.932
24	H	Br	H	H	H	Br	H	H	7.810
25	H	Br	H	H	H	H	H	H	6.530

assays as previously reported.[11,12] The binding affinities are utilized in the CoMFA/QSAR models as the negative log of the effective concentration to displace 50% of radiolabeled ligand in μmol (pEC$_{50}$) and may be found in Tables 1–7.

Molecular Modeling and Alignment Rules. All of the compounds were built, energy minimized, and aligned according to the procedure described by Waller and McKinney.[11,12] Briefly, all molecules were constructed using the Sybyl molecular modeling package and geometry optimized using the standard Tripos molecular mechanics force field using a conjugate gradient minimizer with a 0.001 kcal/mol energy gradient convergence option and a distance-dependent dielectric function (1/r). Atomic charges were determined using the semiempirical package MOPAC[15] with the following keywords: AM1 1SCF NOINTER.

2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD), the prototypical Ah receptor ligand, was used a the template molecule onto which the various dibenzo-p-dioxin, biphenyl, dibenzofuran, naphthalene, and indolocarbazole derivatives were aligned. For the dibenzo-p-dioxin molecules in Table 1, the 2, 3, 7, and 8 carbons of each dibenzo-p-dioxin ring system were aligned using the root-mean-squares (RMS) atom pair fitting algorithm in SYBYL. For the biphenyl congeners in Table 2, the 3, 3′, 4, and 4′ carbons of the biphenyl core were aligned to the 3, 8, 2, and 7 carbons of the template TCDD molecule, respectively, using the same RMS fit alignment procedure. The dibenzofuran molecules in Table 3 were aligned using the same procedure that was implemented for the dibenzo-p-dioxin analogues. The naphthalene

POLARIZABILITY FIELDS FOR USE IN 3D-QSAR

J. Chem. Inf. Comput. Sci., Vol. 41, No. 5, 2001 **1303**

Table 2. Biphenyls Included in CoMFA/QSAR Analyses

no.	R2	R3	R4	R5	R6	R2′	R3′	R4′	R5′	R6′	pEC$_{50}$
26	H	Cl	Cl	H	H	H	Cl	Cl	H	H	6.149
27	H	Cl	Cl	Cl	H	H	H	Cl	H	H	4.553
28	H	Cl	Cl	Cl	H	H	Cl	Cl	H	H	6.886
29	H	Cl	Cl	Cl	H	Cl	H	Cl	H	H	4.845
30	Cl	Cl	Cl	H	H	H	Cl	Cl	H	H	5.367
31	Cl	H	Cl	Cl	H	H	Cl	Cl	H	H	5.041
32	Cl	Cl	Cl	Cl	H	H	H	Cl	H	H	5.387
33	Cl	Cl	Cl	Cl	H	H	Cl	Cl	H	H	5.301
34	Cl	H	Cl	Cl	H	H	Cl	Cl	Cl	H	4.796
35	Cl	Cl	Cl	H	H	H	Cl	Cl	Cl	H	5.149
36	Cl	H	Cl	H	H	Cl	H	Cl	H	H	3.886
37	Cl	H	Cl	Cl	H	Cl	H	Cl	Cl	H	4.102
38	Cl	Cl	Cl	Cl	H	H	H	H	H	H	3.854
39	Cl	H	Cl	H	Cl	H	Cl	Cl	Cl	H	4.004

molecules in Table 4 were aligned in a similar manner by RMS fitting their lateral position carbons (2, 3, 6, and 7) to the corresponding carbons (2, 3, 7, and 8) of TCDD. For the indolocarbazole molecules in Table 5, the lateral carbons of the terminal aromatic rings (A and E) were RMS fitted to the corresponding carbons (2, 3, 7, and 8) of TCDD. The tricyclic cores of the molecules in Tables 6 and 7 were directly superimposed via RMS fit on the tricyclic ring system of TCDD. The resulting alignment rule is graphically depicted in Figure 1. See Tables 1−7 for IUPAC-based numbering of these rings systems. The structures are colored as follows: biphenyl (compound 28) blue, naphthalene (compound 80) green, dibenzofuran (compound 63) orange, dibenzodioxin (compound 1) red, and indolocarbazole (compound 84) yellow.

CoMFA Interaction Energy Calculations. The steric, in terms of van der Waals (6−12) interactions, and electrostatic, Coulombic with a 1/*r* distance-dependent dielectric, potential energy fields were calculated at each lattice intersection on a regularly spaced grid. The grid spacing was either 1.0 or 2.0 Å units in every direction with the grid extending 4.0 Å units in every direction away from the molecule. An sp^3 carbon probe atom with a van der Waals radius of 1.53 Å units and a +1.0 charge was used in all the calculations. The steric energy values were truncated to ±30 kcal/mol.

Calculation of the Polarizability Force Field. Atomic polarizability coefficients are calculated for each molecule using a semiempirical method by the program HSCF,[16] keywords: DE*ORTHO*, FULLDIPOLE, POLARS, EIGENS, *ORTHO*CHARGES. The average value of the polarizability components was read from the output file for each atom of each molecule, this is the atomic polarizability value, α. In a manner analogous to the CoMFA algorithm, the polarizability fields were calculated at each lattice intersection on a regularly spaced grid. The grid spacing was 1.0 or 2.0 Å units in every direction with the grid extending 4.0 Å units
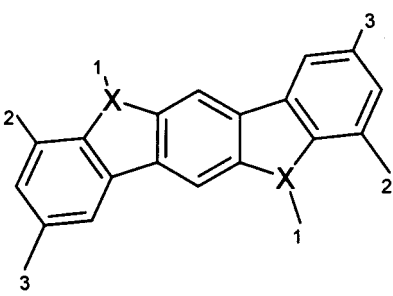
Table 3. Dibenzofurans Included in CoMFA/QSAR Analyses

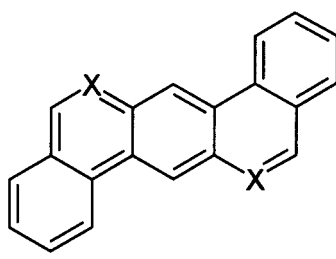| no.	R1	R2	R3	R4	R6	R7	R8	R9	pEC$_{50}$
40	H	Cl	H	H	H	H	H	H	4.061
41	H	H	Cl	H	H	H	H	H	5.003
42	H	H	H	Cl	H	H	H	H	3.429
43	H	Cl	Cl	H	H	H	H	H	6.088
44	H	Cl	H	H	Cl	H	H	H	4.125
45	H	Cl	H	H	H	H	Cl	H	4.103
46	Cl	H	Cl	H	Cl	H	H	H	6.123
47	Cl	H	Cl	H	H	H	Cl	H	4.653
48	H	Cl	Cl	Cl	H	H	H	H	5.396
49	H	Cl	Cl	H	H	H	Cl	H	6.858
50	H	Cl	H	H	Cl	Cl	H	H	7.255
51	H	Cl	Cl	Cl	Cl	H	H	H	7.379
52	H	Cl	Cl	Cl	H	H	Cl	H	7.657
53	Cl	H	Cl	H	Cl	H	Cl	H	7.610
54	H	Cl	Cl	H	H	Cl	Cl	H	8.444
55	Cl	Cl	H	Cl	H	H	Cl	H	5.715
56	Cl	Cl	H	Cl	Cl	Cl	H	H	8.194
57	Cl	Cl	H	Cl	H	Cl	H	Cl	5.371
58	Cl	Cl	Cl	Cl	H	Cl	Cl	H	7.911
59	Cl	Cl	Cl	H	H	Cl	Cl	H	8.147
60	Cl	Cl	H	Cl	H	Cl	Cl	H	6.728
61	H	Cl	Cl	Cl	H	Cl	Cl	H	8.943
62	Cl	Cl	Cl	Cl	H	Cl	Cl	H	7.587
63	Cl	Cl	Cl	H	Cl	Cl	Cl	H	7.508
64	Cl	Cl	H	Cl	Cl	Cl	Cl	H	5.808
65	H	Cl	Cl	Cl	Cl	Cl	Cl	H	8.376
66	H	Cl	Cl	H	Cl	H	Cl	H	7.610
67	Cl	Cl	Cl	H	Cl	H	H	H	7.379
68	Cl	Cl	Cl	H	H	Cl	H	H	7.954
69	Cl	H	Cl	Cl	H	Cl	Cl	H	7.657
70	H	Cl	Cl	Cl	H	Cl	H	Cl	7.657
71	Cl	Cl	Cl	H	H	Cl	H	Cl	7.313
72	H	H	H	H	H	H	H	H	3.429
73	H	Cl	Cl	Cl	H	Cl	H	H	8.689
74	Cl	Cl	Cl	H	H	Cl	H	H	7.954
75	Cl	H	Cl	Cl	H	Cl	Cl	H	7.623
76	H	Cl	Cl	Cl	H	Cl	H	Cl	7.623
77	Cl	Cl	Cl	H	H	Cl	H	Cl	7.313
78	Cl	Cl	H	Cl	Cl	H	Cl	H	6.297

Table 4. Naphthalenes Included in CoMFA/QSAR Analyses

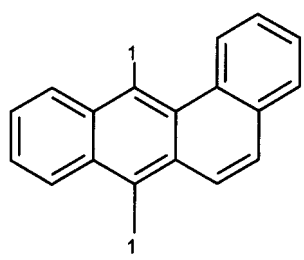| no.	R1	R2	R3	R4	R5	R6	R7	R8	pEC$_{50}$
79	H	Br	Br	H	H	H	H	H	5.616
80	H	Br	Br	H	H	Br	Br	H	7.668
81	Br	Br	H	Br	H	Br	Br	H	7.465
82	Br	Br	Br	Br	H	Br	Br	H	7.608
83	Br	Br	Br	H	Br	Br	Br	H	7.996

in every direction away from the molecule. The grid and molecular alignments used were identical to those used for the CoMFA analyses. The polarizability fields are calculated

Table 5. Indolocarbazoles Included in CoMFA/QSAR Analyses

no.	X	R1	R2	R3	pEC$_{50}$
84	N	CH$_3$	H	H	8.921
85	S	H	H	H	8.482
86	N	H	H	H	8.444
87	N	CH$_2$CH$_3$	H	H	8.051
88	N	COCH$_3$	H	H	7.951
89	N	H	CH$_3$	H	7.721
90	O	H	H	H	7.538
91	N	(CH$_2$)$_3$CH$_3$	H	H	6.824
92	N	H	H	CH$_3$	5.824
93	N	(CH$_2$)$_2$N(CH$_3$)$_2$	H	H	5.824
94	N	H	=O	H	6.824
95	N	H	OH	H	6.824

Table 6. Indolocarbazole Derivatives Included in CoMFA/QSAR Analyses

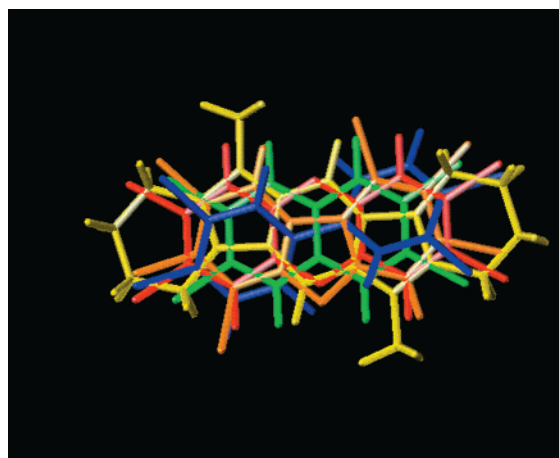| no.	X	pEC$_{50}$
96	C	8.602
97	N	6.863

Table 7. Indolocarbazole Derivatives Included in CoMFA/QSAR Analyses

no.	R1	pEC$_{50}$
98	H	7.319
99	CH$_3$	6.857

for each molecule according to eq 1

$$\forall d_{ij} < vdw_i \!:\! p = \sum_{j=0}^{m} \sum_{i=0}^{n} \alpha_i / d_{ij}^2 \qquad (1)$$

where p is the polarizability (in Å), α is the atomic polarizability coefficient (in Å^3), d is the distance between

Figure 1. Alignment rule.

the grid point and an atom in the molecule (in Å), and *vdw* is the van der Waals radius of an atom in the molecule. As polarizability is arguably an electrostatic phenomenon, the choice of the distance dependency, $1/d_{ij}^2$ in eq 1, is consistent with the Coulombic equation for calculation of the electrostatic field. In preliminary studies (data not shown), $1/d_{ij}^3$ and $1/e^{-d}$ were also evaluated as distance functions in the polarizability equation. The statistical results of QSAR models derived from these fields were not significantly different; however, it was determined that the contour plots were best defined in those models that utilized $1/d_{ij}^2$.

A program written in C is utilized to read in the molecule information from the Sybyl.mol2 file, the polarizability information from the HSCF output file, and the parameters of the CoMFA region from the .rgn file. The output from this program is a file containing the resulting values of the polarizability field at each grid point in the input region computed using eq 1. This field is imported into the Sybyl molecular spreadsheet (MSS) as a new electrostatic type CoMFA column. A script written in Sybyl programming language (SPL) that encompasses all of these steps is used to facilitate the creation of the field is included as Chart 1.

Partial Least Squares (PLS) Analysis. Initially, the PLS algorithm was used in conjunction with the crossvalidation option to obtain the optimal number of components to be used in the subsequent analyses of the data table. The crossvalidated r^2 (termed q^2 herein) was computed as follows

$$q^2 = (\text{SD} - \text{``PRESS''})/\text{SD} \qquad (2)$$

where SD is the sum of the squared deviations between the measured and mean binding affinities of the training set molecules, and "PRESS" is the sum of the squared deviations between the predicted and measured binding affinities for every molecule. For consistency, the leave-one-out implementation of the crossvalidation algorithm was used for all studies.

To increase the signal-to-noise ratio in the CoMFA analyses, individual steric and electrostatic columns corresponding to particular grid intersections exhibiting a standard deviation of less than 0.5 kcal/mol (termed minimum sigma) were not included in the crossvalidated analysis. A minimum sigma value of 0.2 kcal/mol was utilized the analyses of the polarizability fields. Only those columns displaying significant variability in the interaction values measured for

POLARIZABILITY FIELDS FOR USE IN 3D-QSAR

J. Chem. Inf. Comput. Sci., Vol. 41, No. 5, 2001 **1305**

Table 8. Summary Statistics: Diverse Set

										contribution		
n	grid size	fields	σ	q^2	comps	r^2	F	p	se	steric	electro	polar
99	2	st, el	0.5	0.45	7	0.82	61.06	<0.005	0.68	0.49	0.51	
99	1	st, el	0.5	0.65	8	0.91	109.71	<0.005	0.47	0.51	0.49	
99	1	steric	0.5	0.66	6	0.85	89.04	<0.005	0.58	1.00		
99	1	elec	0.5	0.67	8	0.90	109.71	<0.005	0.47		1.00	
99	1	pol	0.5	0.61	5	0.83	92.87	<0.005	0.62			1.00
99	1	pol, st	0.5	0.66	7	0.90	116.67	<0.005	0.49	0.51		0.49
99	1	pol, el	0.5	0.69	7	0.92	148.58	<0.005	0.43		0.64	0.36
99	1	pol, st, el	0.5	0.72	6	0.91	161.13	<0.005	0.45	0.26	0.48	0.26

Chart 1

```
uims define macro polarize Sybylbasic yes


setvar reg %prompt(string "reg" "Enter name of region file:")
setvar column %prompt(string "column" "Enter number of new comfa column:")

# Calculate the polarizability field file for each molecule
# and import the field into the new comfa column.

setvar count 1

setvar dbname %table_attribute(database)
setvar dbname %cat($dbname /)
for i in %table(* row name)
   setvar molname %lowercase($i)
   setvar fieldfile %cat($molname .pfu)
   setvar fieldfile %cat($dbname $fieldfile)
   setvar fielddata %cat($molname .pf)
   setvar fielddata %cat($dbname $fielddata)
   setvar polarname %cat($molname .hscf)
   setvar molname %cat($dbname $molname)
   setvar molfile %cat($molname .mol2)
   setvar polarfile %cat($dbname $polarname)
   dcl hscf $molfile mol2 $polarfile DEORTHO FULLDIPOLE POLARS EIGENS
ORTHOCHARGES
   dcl new_polar $molname $reg mo

   qsa com field import $reg $fieldfile $fielddata ||
   qsa tab enter cell $count $column yes $fieldfile ||

   setvar count %math($count+1)

endfor

#Evaluate the new column

qsa table evaluate new * $column
```

different molecules were considered. This "column filtering" technique typically reduces the number of columns in the QSAR table by a factor of 10. The PLS analysis was then repeated without crossvalidation, and the number of components designated as that which yielded the highest q^2 value, to generate conventional r^2 values. For the non-crossvalidated analyses, the minimum sigma was reduced to 0.0 kcal/mol (all columns were included in the analyses).

A CoMFA was performed according to the parameters specified in the original analyses[12] (i.e. steric and electrostatic field parameters and 2.0 Å grid spacing as reported) in order to reproduce the results from the previous analyses and to form a basis for comparison with the new field results. The CoMFA was then repeated using a 1.0 Å grid spacing and identical energy cutoff parameters for the steric and electrostatic fields. Separate CoMFA models were constructed based on the steric field, the electrostatic field, the polarizability field, and all possible combinations of two and three fields calculated on the 1.0 Å spaced region.

RESULTS AND DISCUSSION:

Statistical Results. The statistical results for all combinations of fields in the 3D-QSAR models for the diverse molecules are summarized in Table 8. The analysis using standard CoMFA steric and electrostatic fields and a default grid spacing of 2.0 Å yielded a model that was internally

consistent ($q^2 = 0.45$ using seven principle components, PCs) and statistically robust ($r^2 = 0.82$). The steric and electrostatic fields contributed nearly equally to the model at 49% and 51%, respectively. The standard steric and electrostatic fields CoMFA model calculated on a region with a 1.0 Å grid spacing exhibited superior results to the model derived from fields calculated on a 2.0 Å spacing ($q^2 = 0.65$ and $r^2 = 0.91$ using eight PCs). The single steric and electrostatic field based model yielded results that were not statistically different from those of the standard two-field model. The polarizability field produced a model with results that were very comparable to those of the steric and electrostatic field-based models in terms of internal predictive ability ($q^2 = 0.61$ using five PCs) and statistical validity ($r^2 = 0.83$). The polar field, when used in combination with the steric field, contributed almost equivalently to the model yielding results very similar to those above. It is interesting to note that the polar field contributed somewhat less, approximately 36%, to the model when used in combination with the electrostatic field producing a model with a $q^2 = 0.69$ using seven PCs and an $r^2 = 0.92$. In the three-field model, the electrostatic field was the dominant field displaying a contribution level of 48%, while the steric and polarizability fields contributed 26% each to the model. This model yielded the best results in terms of internal predictability and statistical significance ($q^2 = 0.72$ using six PCs and $r^2 = 0.91$). In comparison, the polarizability field only model derived from five principle components displayed a nearly equivalent level of internal consistency and was only slightly less robust than the three-field model that required eight principle components. These results support the use of polarizability fields as predictive fields in 3D-QSAR studies. A predicted versus actual plot from the polarizability field only model is shown in Figure 2.

Field Correlation Analysis. Since the polarizability field, in theory, embodies characteristics of both steric and electrostatic fields, a field correlation analysis was performed in order to ascertain the potential degree of redundancy between the fields. An initial analysis of the steric and electrostatic fields revealed an average correlation (as r) of -0.11. The electrostatic and polar fields were determined to be less correlated expressing an average r of -0.04. The greatest, yet statistically insignificant, correlation was identified as being between the steric and polarizability fields with an average r of 0.27.

Steric Contour Plots. The results of CoMFA studies are best viewed as three-dimensional color-coded contour plots. A contour plot of the stddev*coeff values for the steric field from the three-field CoMFA is presented in Figure 3 with molecule 89, a indolocarbazole derivative, shown for refer-
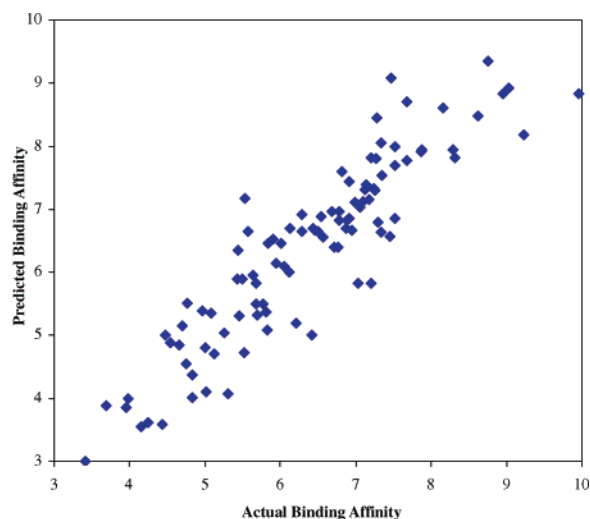
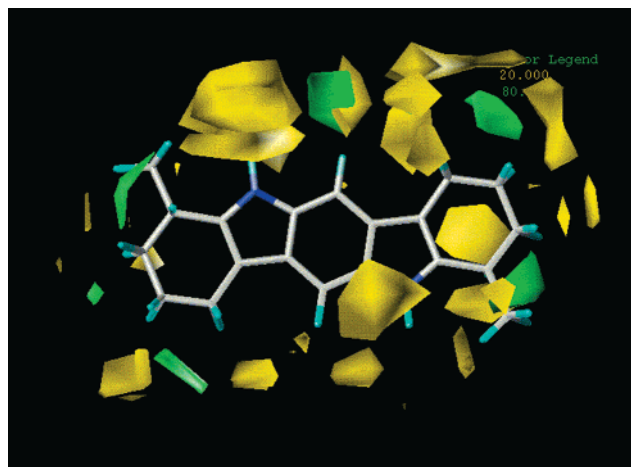**Figure 2.** Predicted versus actual binding affinity values from polarizability QSAR model.

Figure 3. View of CoMFA steric contour plots for data set 3 (diverse). Contours are stddev*coeff and plotted as contribution. Regions of positive contribution (greater bulk desired) are shown in green (80% contribution) and regions of negative steric contribution (less bulk desired) are shown in yellow (20% contribution). Molecule 89 is shown for reference.

ence. Green polyhedra represent areas in the pseudoreceptor space around the aligned molecules that are tolerant to steric bulk. These are found primarily in the so-called lateral regions relative to the molecular core. These regions correspond to the 2, 3, 7, and 8 positions on the dibenzo-p-dioxin and dibenzofuran cores, the 2, 3, 6, and 7 positions on the naphthalenes, and the 3 (or 3′) and 4 (or 4′) positions of the biphenyl ring systems. In Figure 3, it is apparent that these areas of steric bulk tolerance are near the methyl substitituent on the indolocarbazole core. These results suggest that regardless of the molecular core, a certain amount of steric bulk is desired in these lateral areas.

In Figure 3, yellow polyhedra denote areas in the pseudoreceptor space that are intolerant to steric bulk enhancements. Most obvious in this figure are the yellow polyhedra that are located in the vicinity of the nitrogens in the carbazole. Substituents at this position correspond to substitutions on the dibenzo-p-dioxin and dibenzofuran cores at the 1, 4, 6, and 9 positions, the 1, 4, 5, and 8 positions on the naphthalenes, and at the ortho positions on the biphenyl molecules included in the training set.
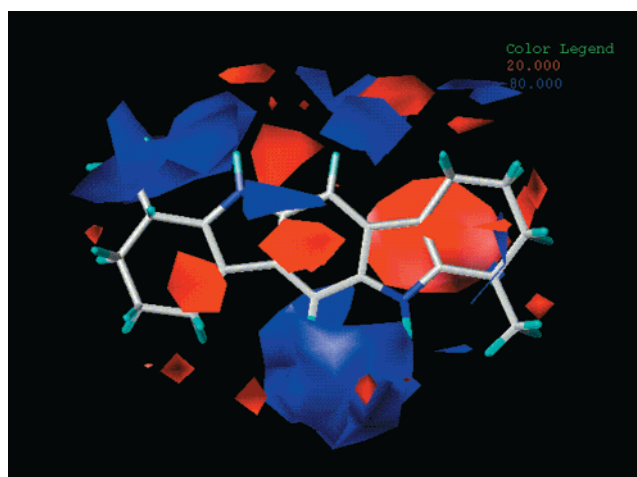

Figure 4. View of CoMFA electrostatic contour plots for data set 3 (diverse). Contours are stddev*coeff and plotted as contribution. Regions of positive contribution (greater positive charge desired) are shown in blue (80% contribution), and regions of negative contribution (less positive charge desired) are shown in red (20% contribution). Molecule 89 is shown for reference.

Electrostatic Contour Plots. The corresponding plot for the electrostatic field from the three-field CoMFA is found in Figure 4. These plots are, in general, more difficult to interpret than steric contour plots. However, it is suggested by the blue polyhedra that increased positive charge is desired in the vicinity of the medial substituents on all molecular cores. This indicates that, for this diverse data set, relatively active compounds are found that are not halogenated in these positions. Additionally, these positions correspond to the ortho positions on the biphenyl derivatives suggesting that halogenation is not desired as these positions. A large region of positive charge tolerance is also noted near the methyl substituent on the reference molecule, 89. This suggests that compounds need not be fully halogenated in the lateral positions in order to be effective ligands for the Ah receptor. An analysis of the biological data set for the dibenzo-p-dioxins (Table 1) confirms this.

In Figure 4, regions of negative electrostatic potential are noted in the vicinity of the lateral substituents on the phenyl ring systems of the dibenzo-p-dioxin, dibenzofuran, naphthalene, and biphenyl molecules included in the training set. This is consistent with the optimal halogenation pattern on these types of molecules. It is interesting to note that the terminal aromatic rings of the carbazole shown in Figure 3 correspond to these substituents.

For a more detailed interpretation of this analysis, the reader is referred to the original publication of this model.[12] These results provide additional insight into the known structure−activity relationships and indicate that the CoMFA is extracting the appropriate information from the potential energy matrix, thus providing a measure of positive control.

Polarizability Contour Plots. In Figure 5, the polarizability field from the analysis is plotted. As was the case for the electrostatic contour plot, the diverse nature of the training set makes this plot somewhat difficult to interpret. Yet, it is possible to draw some general conclusions. In the figure, blue polyhedra are used to denote areas around the aligned molecular cores that desire enhanced polarizability. These regions are seen near the lateral positions of the various molecular cores. For the majority of the compounds in the
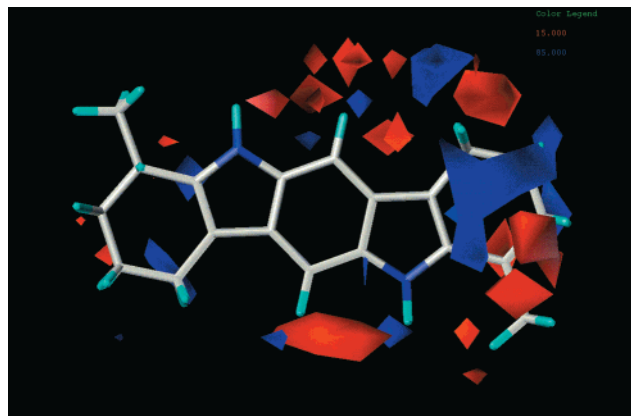
POLARIZABILITY FIELDS FOR USE IN 3D-QSAR

J. Chem. Inf. Comput. Sci., Vol. 41, No. 5, 2001 **1307**

Figure 5. View of CoMFA polarizability contour plots for data set 2 (diverse). Contours are stddev*coeff and plotted as contribution. Regions of positive contribution (greater polarizability desired) are shown in blue (85% contribution), and regions of negative contribution (less polarizability desired) are shown in red (15% contribution). Molecule 89 is shown for reference.

training set, halogenation is required at these positions in order to facilitate binding. For the indolocarbazole derivatives, it is hypothesized that the aromatic nature of the lateral phenyl rings acts as a surrogate for the electron rich halogens. This result is very consistent with our existing knowledge of the structure−activity relationships for these molecules in that it has been previously hypothesized that increasing the polarizability along the long axis of the molecule is beneficial for binding.[9] Regions of space that require a corresponding decrease in the polarizability are found closer to the medial positions and are denoted by red polyhedra.

CONCLUSIONS

The theory and application of molecular polarizability fields in 3D-QSAR studies has been presented. For the data set studied, in which polarizability has previously been determined to be an important parameter in binding, the polarizability fields yielded models that were statistically equivalent to standard CoMFA steric and electrostatic fields. While it is true that the steric field contains information that includes a crude approximation of molecular polarizability through dispersion forces, a field correlation analysis revealed that the steric and polarizability fields were providing different types of information. When polarizability is a dominant factor in determining the strength of the interaction between molecules, as is the case for both data sets under examination, the polarizability component of the steric interactions may be too crude to provide interpretable results.

While it is expected that polarizability fields will be of use in studies in which receptor:ligand binding events involve charge-transfer or π-stacking interactions, future studies will also focus on the applicability and utility of these fields in model systems for the analysis of the molecular recognition event in protein:protein interactions. The possibility that polarizability fields could provide additional insight into the mechanics of membrane permeability as it relates to drug absorption is also currently being explored.

REFERENCES AND NOTES

(1) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(2) Cramer, R. D.; Bunce, J. D.; Patterson, D. E.; Frank, I. E. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18−25.
(3) Kearsley, S. K.; Smith, G. M. *Tetrahedron Comput. Met.* **1990**, *3*, 615−633.
(4) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. *J. Comput.-Aided Mol. Des.* **1991**, *5*(6), 545−552.
(5) Boobbyer, D. N.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. *J. Med. Chem.* **1989**, *32*(5), 1083−1094.
(6) Waller, C. L.; Marshall, G. R. *J. Med. Chem.* **1993**, *36*, 2390−2403.
(7) Waller, C. L.; Evans, M. V.; McKinney, J. D. *Drug Metab. Disp.* **1996**, *24*(2), 203-210.
(8) Safe, S. H. *Annu. Rev. Pharmacol. Toxicol.* **1986**, *26*, 371−399.
(9) McKinney, J. D.; Gottschalk, K. E.; Pedersen, L. *J. Mol. Struct.* **1983**, *105*, 427−438.
(10) Applequist, J.; Carl, J. R.; Fung, K.-K. *J. Am. Chem. Soc.* **1979**, *101*, 2952.
(11) Waller, C. L.; McKinney, J. D. *J. Med. Chem.* **1992**, *35*, 3660−3666, and references therein.
(12) Waller, C. L.; McKinney, J. D. *Chem. Res. Toxicol.* **1995**, *8*, 847−858, and references therein.
(13) SYBYL 6.5; Tripos, Inc.: St. Louis, MO.
(14) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. *SIAM: J. Sci. Stat. Comp.* **1984**, 735−743.
(15) MOPAC, Quantum Chemistry Program Exchange, #455, Indiana University, IN.
(16) HSCF, Prof. R. Perlman, The University of Texas, Austin, TX.

CI0004659