

QSAR's from Similarity Matrices. Technique Validation and Application in the Comparison of Different Similarity Evaluation Methods

Andrew C. Good,[†] Stephen J. Peterson, and W. Graham Richards*

Physical Chemistry Laboratory, South Parks Road, Oxford, United Kingdom

Received February 10, 1993*

It has recently been shown that good quantitative structure-activity relationships can be obtained through statistical analysis of molecular similarity matrices. Here we extend the technique to seven additional molecular series, previously studied using Comparative Molecular Field Analysis (CoMFA) methodology. The results are used to confirm technique applicability across a wider range of QSAR problems and to compare quantitatively the ability of various similarity indices to describe biological systems. The relative merits of this technique in comparison to CoMFA are discussed.

Introduction

Numerical indices which measure the overall electrostatic and steric similarity between pairs of molecules have been used for some time in activity data correlations.¹⁻³ In these applications, correlation was achieved by relating activity to the similarity data obtained from comparisons to a single lead molecule. Recently⁴ an alternative approach was tested in which the data matrix obtained from a full N by N (each molecule compared to every other) similarity calculation was analysed using partial least squares (PLS) statistics.⁵ Analysis of the full matrix implicitly introduces some of the location dependence of the steric and electrostatic parameters used within CoMFA methodology,⁶ and was found to give excellent correlation with binding data for a steroid data set. In this study we have extended the approach to cover seven additional molecular series.

The aim of these investigations is 2-fold. While the results for the steroid data set were very encouraging, we wanted to show that the technique could be applied to a wide variety of different problems. We were also interested in applying the methodology to gain a better insight into similarity index utility. Up to now index quality has generally been evaluated through comparison of structure ranking for similarity calculations against a lead molecule.^{3,7-9} While this gives qualitative information regarding the general comparative behavior of the various functions, it would be preferable if one were able to gain a more quantitative insight. By calculating the correlation coefficient for the similarity matrix of a given index and data set, one can obtain a quantitative measure of the ability of that index to describe biological data. Since it is the aim of a drug-design project to describe the biological system under investigation, a measure of this kind would seem a good way to evaluate function and index utility.

The following indices have been tested using this technique.

Carbo Index^{10,11}

Probably the most widely applied formula for the calculation of molecular similarity is the Carbo index.

$$R_{AB} = \frac{\int P_A P_B d\nu}{(\int P_A^2 d\nu)^{1/2} (\int P_B^2 d\nu)^{1/2}}$$

Molecular similarity R_{AB} is determined from the structural properties P_A and P_B of the two molecules being compared. The numerator measures property overlap while the denominator normalizes the similarity result. As originally applied, electron density was used as the structural property P . For this investigation, however, electrostatic potential, electric field, and shape have been used through application of modified versions of the ASP program.¹²

Electrostatic potential (and electric field where applicable) for this and all other indices have been evaluated at the intersections of a rectilinear grid constructed around the two molecules using point charge data. In order to avoid singularities at the atomic nuclei (where $1/r$ tends to infinity), evaluation is restricted to points outside the van der Waals volume of the molecules in the calculation. The resulting electrostatic values are then used to evaluate the indices numerically.

Shape is evaluated in a similar manner using a modified version of the Carbo equation proposed by Meyer.¹³ The mechanics of the similarity evaluation are the same as those applied to electrostatic potential and electric field calculations. For shape, every grid point is tested to see whether it falls inside the van der Waals surface of each molecule. The results are then applied to the following modified version of the Carbo index

$$S_{AB} = \frac{B}{(T_A T_B)^{1/2}}$$

B is the number of grid points falling inside both molecules, while T_A and T_B are the total number of grid points falling inside each individual molecule.

For the Carbo index an alternative analytical approach to evaluating electrostatic and shape similarity was also tested. Here the numerical grid evaluation was replaced by analytical Gaussian function calculations. For elec-

[†] Present address: School of Pharmacy, Department of Pharmaceutical Chemistry, University of California, San Francisco.

* Abstract published in *Advance ACS Abstracts*, September 1, 1993.

trostatic potential, one, two, and three Gaussian function approximations to the inverse distance dependence of point charge electrostatic calculations were used.⁷ For shape Gaussian functions determined from the square of the STO-3G atomic orbital¹⁴ wave functions were used.⁸ Three different sets of "shape" Gaussian functions were considered.

(1) Three Gaussians were fitted directly to the STO-3G atomic orbital derived electron density of each atom type (henceforth known as "unmodified Gaussians").

(2) To account for atoms appearing "harder" within molecules, the three Gaussians were also fitted to a modified version of the electron density function (henceforth referred to as "VDW fitted Gaussians"). For this modified function, the electron density was set to zero beyond the van der Waals radius of each atom type.

(3) Hydrogen atoms are particularly soft; i.e. the electron density extends well beyond the van der Waals radius. As a consequence it was felt that use of the modified Gaussians might only be required for the hydrogen atom type. A third set of functions were therefore created using the VDW fitted Gaussians for hydrogen and the unmodified Gaussians for all other atom types (henceforth known as "H VDW fitted Gaussians").

Hodgkin Index¹⁵

The Carbo index is sensitive to the shape of a property's distribution rather than to its magnitude. This is highlighted by the fact that when the electron densities of two molecules correlate, the similarity index tends toward unity. Thus if $P_A = nP_B$, R_{AB} equals unity. To increase the sensitivity of the formula to a property's magnitude, the Hodgkin index was created.

$$H_{AB} = \frac{2 \int P_A P_B d\nu}{\int P_A^2 d\nu + \int P_B^2 d\nu}$$

This index has the effect that, if $P_A = nP_B$, then $H_{AB} = 2n/(1 + n^2)$.

Electrostatic potential and electric field similarity were analyzed using this index.

Linear and Exponential Indices¹⁶

These indices evaluate electrostatic similarities at each individual grid point, the values are combined, and the total is divided by the number of grid points involved to determine the average similarity.

$$L_{AB} = \frac{\sum_{i=1}^n (1 - X)}{n} \quad E_{AB} = \frac{\sum_{i=1}^n \exp^{-X}}{n}$$

where

$$X = \frac{|P_A - P_B|}{\max(|P_A|, |P_B|)}$$

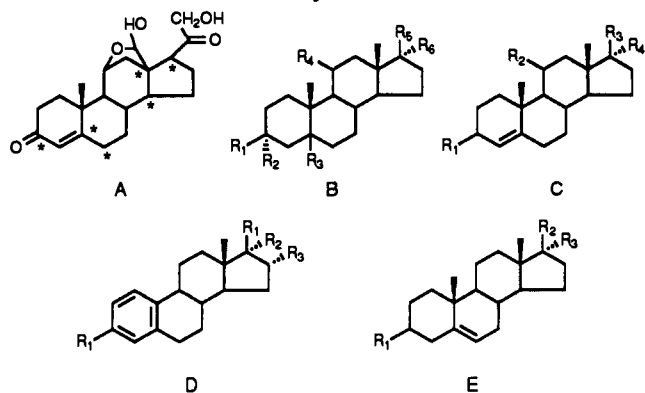
$\max(|P_A|, |P_B|)$ (P_{\max}) is the larger electrostatic potential magnitude between P_A and P_B at the grid point where the similarity is being calculated. n is the number of grid points involved in the calculation.

Electrostatic potential similarity has been evaluated using these indices.

Spearman Rank Correlation Coefficient¹⁷

This versatile formula has previously been used to measure the electrostatic potential similarity of two

Table I. Structure and Affinity Data for the Steroids of Series a



no.	structure	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	(i) CBG log 1/K	(ii) TBG log 1/K
1	A							-6.279	-5.322
2	B	OH	H	H	H	OH	H	-5.000	-9.114
3	E	OH	OH	H				5.000	-9.176
4	C	=O	H	=O				-5.763	-7.462
5	B	OH	H	H	H	=O		-5.613	-7.146
6	C	=O	OH	COCH ₂ OH	H			-7.881	-6.342
7	C	=O	OH	COCH ₂ OH	OH			-7.881	-6.204
8	C	=O	=O	COCH ₂ OH	OH			-6.892	-6.431
9	E	OH	=O					-5.000	-7.819
10	C	=O	H	COCH ₂ OH	H			-7.653	-7.380
11	C	=O	H	COCH ₂ OH	OH			-7.881	-7.204
12	B	=O		H	H	OH	H	-5.919	-9.740
13	D	OH	H	H				-5.000	-8.833
14	D	H	OH	OH				-5.000	-6.633
15	D	=O		H				-5.000	-8.176
16	B	H	OH	H	H	=O		-5.225	-6.146
17	E	OH	COCH ₃	H				-5.225	-7.146
18	E	OH	COCH ₃	OH				-5.000	-6.362
19	C	=O	H	COCH ₃	H			-7.380	-6.944
20	C	=O	H	COCH ₃	OH			-7.740	-6.996
21	C	OH	H	OH	H			-6.724	-9.204

molecules over the intersections of a rectilinear grid.

$$R_{AB} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}$$

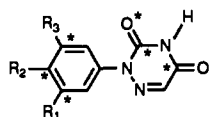
d_i is the difference in the electrostatic potential rank at point i of two structures, and n is the total number of grid points.

Electrostatic potential similarity has been calculated using this index.

Experimental Section

To test index utility and general technique applicability, the following series were used: (a) 21 steroids with binding data to corticosteroid and testosterone binding globulins (Table I),⁴ (b) 54 anticoccidial triazines (Table II),¹⁸ (c) 37 β -carboline, pyridodiindole, and CGS ligands with binding data for the benzodiazepine receptor inverse agonist site (Table III),¹⁹ (d) 14 β -(*p*-substituted phenyl)tropane-2 β -carboxylic acid methyl esters with affinity data to the cocaine binding site (Table IV),²⁰ (e) MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) and 40 analogues with qualitative affinity data indicators for binding to monoamine oxidase (Table V),²¹ (f) 28 clonidine-like imidazole analogues with pK_a data (Table VII),²² (g) 16 2-substituted imidazoles with pK_a data (Table VIII),²² (h) 49 substituted benzoic acids with Hammett constant data (Table VI).²³

Model Building. Series a (steroids) and d (cocaine analogues) were constructed from the co-ordinates of related structures in the Cambridge Crystallographic Database.²⁴ Any additional functional groups were added using the modeling program CHEMX²⁵ as required, ensuring that the starting geometry of any given substituent was identical. The remaining series were

Table II. Structure and Activity Data for the Triazines of Series b

no.	R ₁	R ₂	R ₃	log 1/MEC
1	Cl	CH(CN)C ₆ H ₄ -4-Cl	Cl	3.61
2	Cl	SC ₆ H ₄ -4-Cl	Cl	3.30
3	Cl	SC ₆ H ₄ -4-Cl	CH ₃	3.28
4	Cl	SC ₆ H ₄ -4-COCH ₃	CH ₃	3.19
5	Cl	CH ₂ C ₆ H ₄ -4-Cl	Cl	3.18
6	Cl	SO ₂ C ₃ H ₄ -4-Cl	Cl	2.94
7	Cl	COC ₆ H ₄ -4-Cl	Cl	2.90
8	CH ₃	SC ₆ H ₄ -4-Cl	CH ₃	2.86
9	CH ₃	SO ₂ C ₃ H ₄ -4-Cl	CH ₃	2.59
10	CH ₃	S(O)C ₆ H ₄ -4-Cl	CH ₃	2.58
11	CH ₃	OC ₆ H ₃ -3-CH ₃ -4-SCH ₃	CH ₃	2.27
12	Cl	OC ₆ H ₄ -4-COCH ₃	CH ₃	2.27
13	Cl	OC ₆ H ₄ -4-CH(OH)CH ₃	CH ₃	2.27
14	Cl	SC ₆ H ₄ -4-Cl	H	2.26
15	Cl	CH(OH)C ₆ H ₄ -4-Cl	H	2.26
16	CH ₃	OC ₆ H ₄ -4-SCH ₃	CH ₃	2.25
17	CH ₃	OC ₆ H ₄ -4-SCH ₃	H	2.23
18	Cl	OC ₆ H ₃ -2-Cl-4-SO ₂ NCH ₃ C ₂ H ₅	CH ₃	2.11
19	Cl	OC ₆ H ₄ -4-Cl	CH ₃	1.98
20	Cl	OC ₆ H ₄ -4-SCH ₃	H	1.98
21	CH ₃	OC ₆ H ₄ -4-SO ₂ CH ₃	H	1.97
22	H	SO ₂ C ₆ H ₄ -4-Cl	H	1.96
23	Cl	H	Cl	1.81
24	Cl	OC ₆ H ₄ -4-I	H	1.77
25	Cl	O(naphth-2-yl)-6-Br	H	1.77
26	H	SO ₂ C ₆ H ₄ -4-Br	H	1.71
27	Cl	OC ₆ H ₃ -2,4-Cl ₂	H	1.71
28	Cl	SO ₂ C ₃ H ₄ -4-Cl	CH ₃	1.71
29	CH ₃	OC ₆ H ₄ -4-Br	H	1.70
30	CH ₃	OC ₆ H ₃ -2,4-Cl ₂	H	1.69
31	Cl	CH ₂ C ₆ H ₄ -4-Cl	H	1.64
32	Cl	OC ₆ H ₃ -2-Clh4-SO ₂ NH-c-C ₃ H ₅	CH ₃	1.51
33	CH ₃	SO ₂ N(CH ₂ CH ₂) ₂ O	CH ₃	1.39
34	CH ₃	COC ₆ H ₄ -4-Cl	CH ₃	1.38
35	CF ₃	Br	H	1.35
36	CF ₃	F	H	1.26
37	Cl	H	CH ₃	1.20
28	Cl	SO ₂ N(CH ₂ CH ₂) ₂ O	Cl	1.13
39	Cl	SO ₂ N(CH ₂ CH ₂) ₂ O	H	1.09
40	H	SO ₂ C ₃ H ₅	H	1.04
41	CH ₃	CH(OH)C ₆ H ₅	H	1.01
42	CF ₃	H	H	0.93
43	CH ₃	H	CH ₃	0.86
44	CN	H	H	0.85
45	CH ₃	OC ₆ H ₄ -4-Cl	H	0.74
46	Cl	S(naphth-2-yl)	H	0.48
47	Cl	H	H	0.25
48	Cl	OCH ₃	Cl	0.06
49	H	NO ₂	H	-0.03
50	H	OC ₆ H ₅	H	-0.25
51	H	SO ₂ CH ₃	H	-0.27
52	OCH ₃	H	OCH ₃	-0.30
53	H	H	H	-0.42
54	OCH ₃	OCH ₃	OCH ₃	-0.55

built from scratch within CHEMX using standard bond lengths and angles, again ensuring identical starting geometry for any given substituent. Once built, each structure was minimized within CHEMX using the default force field. AM1 MOPAC²⁶ calculations were then undertaken from which point charges were back-calculated to fit the consequent molecular electrostatic potentials using the RATTler^{12,27} software. Finally the resultant structures were superimposed by least-squares fitting. The atoms involved in the fitting for any given series have been starred on the reference structure above each relevant table (equivalent atoms used where multiples structures are shown), or are described in the associated table legend. For series a-d the fitting template was the most active molecule in the series. The MPTP parent structure was used as the template for series e. Series f and g were fitted onto the first molecule of the series.

Similarity Calculations. For all series the following electrostatic potential and electric field N by N similarity calculations were run to text index utility: (i) a Carbo index with numerical grid based integral evaluation ("Carbo" in the table data), (ii) a Carbo index with a three-Gaussian-function approximation to the $1/r$ curve ("3-Gaussian Carbo" in the table data), (iii) a Carbo index with a two-Gaussian-function approximation to the $1/r$ curve ("2-Gaussian Carbo" in the table data), (iv) a Carbo index with a one-Gaussian-function approximation to the $1/r$ curve ("1-Gaussian Carbo" in the table data), (v) a Hodgkin index ("Hodgkin" in the table data), (vi) a linear index ("linear" in the table data), (vii) an exponential index ("exponential" in the table data), and (viii) a Spearman rank correlation coefficient ("Spearman" in the table data).

The following electric field similarity calculations were also run for all series: (i) a Carbo index ("Carbo field" in the table data) and (ii) a Hodgkin index ("Hodgkin field" in the table data).

For all the above grid-based calculations, an increment of 1.0 Å was used in conjunction with an extent of 4.0 Å. A 5 kcal mol⁻¹ electrostatic potential magnitude cutoff was used in determining which grid points were involved in Spearman rank correlation and linear and exponential index calculations.

For series a-e the shape similarity calculations listed below were run. [Note that Gaussian function electron density approximations were only parameterized for carbon, hydrogen, oxygen, nitrogen, and sulfur.⁸ As a consequence of this, structural subsets of series a-e containing only these atom types were used in their statistical analysis. Series f-h were ignored for shape, as correlation was with electronic rather than affinity data.] (i) a Carbo index with unmodified STO-3G Gaussian function electron density approximations ("unmodified Gaussians" in the table data), (ii) a Carbo index with H VDW fitted STO-3G Gaussian function electron density approximations ("H VDW-fitted Gaussians" in the table data), (iii) a Carbo index with VDW fitted STO-3G Gaussian function electron density approximations ("VDW-fitted Gaussians" in the Table data), and (iv) a Carbo index using the Meyer grid based evaluation method for the structural subsets. For this system two runs were made for each series. In the first run only those structures used with the Gaussian functions were processed ("grid based shape reduced set" in the table data). In the second run the full data set was used ("grid based shape full set" in the table data).

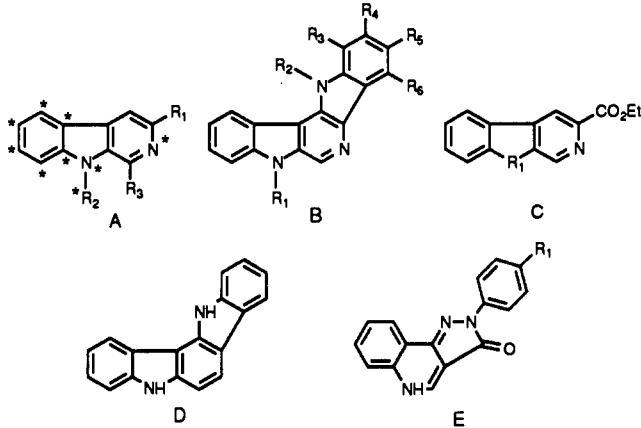
For the above grid based shape similarity calculations an increment of 0.5 Å was used in conjunction with an extent of 10 Å.

A number of additional evaluations were undertaken for the grid-based similarity calculations. For these evaluations varying increments were used to determine how tolerant similarity calculations are to the use of coarser grids.

For all series the Hodgkin index was tested in conjunction with 2.0 and 3.0 Å grid increments using electric field. For series a-e the Meyer's grid-based shaped evaluation was tested at increments of 0.2, 1.0, and 2.0 Å. Results are shown in Table XII.

Statistical Analysis. A symmetric N by N data matrix was output from the modified ASP programs. Column 1 and row 1 contain the similarities of structure 1 to all other compounds, column 2 and row 2 the same for structure 2, etc. The matrix diagonal is made up of a series of 1.0 values, representing the similarity of each structure to itself. The unmodified matrix was analyzed with the PLS module of the GOLPE program²⁸ using leave-one-out cross-validation of all structures over 10 components. The result extracted was the first maximum value in the 10 component list. Matrices with more than 30 columns were then reduced in size through D optimization of their GAMMA values.^{28,29} This procedure keeps only the most orthogonal data for a given number of components, thus reducing the dimensionality of the data matrix while losing the minimum of information. The number of components used was equal to that which yielded the best cross-validated r^2 or 3, whichever was smaller. Half the variables were retained.

The D-optimized matrices and all other matrices with fewer than 30 columns were then reduced using progressive fixing exclusion.²⁸ This process attempts to determine which variables aid correlation through the application of design matrices

Table III. Structure and Affinity Data for the β -Carbolines, Pyridodiindoles, and CGS Compounds of Series c^a


no.	structure	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	log (1/IC ₅₀)
1	A	CO ₂ CH ₃	H	H				-0.699
2	A	CO ₂ CH ₂ CH ₃	H	H				-0.699
3	B	H	H	H	H	H	H	-0.602
4	A	OCH ₂ CH ₃	H	H				-1.380
5	A	OCH(CH ₃) ₂	H	H				-2.699
6	A	OCH ₂ CH ₂ CH ₂ CH ₃	H	H				-1.991
7	A	OCH ₃	H	H				-2.093
8	A	OCH ₂ CH ₂ CH ₃	H	H				-1.042
9	A	COCH ₂ CH ₂ CH ₃	H	H				-0.447
10	A	CH ₂ CH ₂ CH ₂ CH ₃	H	H				-2.389
11	B	H	H	CH ₃	H	H	H	-1.919
12	B	H	H	H	CH ₃	H	H	-1.000
13	B	H	H	H	H	CH ₃	H	-2.350
14	B	H	H	H	H	H	CH ₃	-3.836
15	B	CH ₃	H	H	H	H	H	-3.066
16	B	H	CH ₃	H	H	H	H	-2.196
17	B	CH ₃	CH ₃	H	H	H	H	-3.283
18	E							-3.295
19	A	H	H	H				-3.210
20	A	CO ₂ C(CH ₃) ₃	H	H				-1.00
21	C	C(=O)						-4.415
22	C	C(=NOH)						-3.699
23	C	O						-3.964
24	C	CH ₂						-2.833
25	D	H						0.398
26	D	Cl						0.222
27	DF	OCH ₃						1.000
28	B	H	H	H	H	H	OCH ₃	-2.398
29	B	H	H	H	H	H	Cl	-2.854
30	A	Cl	H	H				-1.653
31	A	NO ₂	H	H				-2.097
32	A	CO ₂ CH ₂ C(CH ₃) ₃	H	H				-2.875
33	A	CO ₂ CH ₃	H	CH ₂ CH ₃				-3.877
34	A	H	H	CH ₂ CH ₃				-5.398
35	A	H	H	CH ₃				-4.093
36	C	C(=O)N(H)						-3.380
37	C	S						-3.230

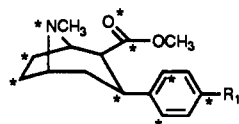
^a R₂ group atom (or its equivalent) α to nitrogen used in rigid fit.

processed using fractional factorial design.³⁰ Large numbers of different data matrices are created in which variables are included or excluded according to the current design cycle. The resultant matrices are used to form a new cross-validated PLS model, and the predictivity of the models is used to determine which variables improve predictive correlation (for a more detailed discussion of these techniques, see ref 28). All variables which are determined to be noise in the system are excluded. GOLPE retains all variables determined to be signal together with those of uncertain property. It is recommended, however, that data of unknown effect are also excluded to maximize the robustness of the model.²⁸ These data were thus also removed manually, leaving only variables proven to aid predictive correlation. Three runs were undertaken for each matrix. For systems with a greater than 1 component maximum cross-validated r^2 , two-, three-, and four-component runs were undertaken. For systems with a 1 component maximum one-, two-, and three-component runs were undertaken. Cross-validation was undertaken using 30 random groups of five molecules from the test set. The ratio of variables

to dummies was set to 1 with a design combination to variables ratio of 5. The final models were tested by repeating the leave-one-out cross-validation analysis for up to three components. The model yielding the highest cross-validated r^2 for three components or less was retained. Results for electronic similarity calculations are shown in Table IX. Results for shape-similarity calculations are shown in Table X. Results for the systems run in variable grid increments are shown in Table XI. Average cross-validated r^2 values for each of the index calculations are shown in Table XII.

Additional calculations were also undertaken using mixed property matrices for series a-e. For each series the reduced matrix of the grid-based full-set shape-similarity calculation was merged with the most predictive reduced electronic data matrix. The data were autoscaled (the variance of shape-similarity data is lower than that of electronic similarity data) and the progressive exclusion procedure repeated. Results for these mixed-matrix calculations are shown in Table XIII. In all cases the results

Table IV. Structure and Affinity Data for the Cocaine Analogues of Series d



no.	R ₁	log (1/IC ₅₀)
2	H	-1.36
3	F	-1.20
4	Cl	-0.07
5	CH ₃	-0.23
6	CF ₃	-1.12
7	CH ₃ O	-0.91
8	NO ₂	-1.00
9	NH ₂	-1.40
10	C ₂ H ₅ CONH	-2.50
11	CH ₃ CONH	-1.81
12	C ₂ H ₅ CONH	-1.83
13	Br	-0.26
14	I	-0.10
15	N ₃	-0.33

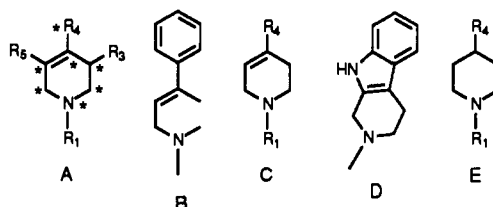
obtained using the unmodified and GOLPE-reduced matrices have been included.

Discussion

The statistical data shown in Tables IX–XIII highlight the utility of the GOLPE package. Reduced-matrix models invariably produced improved cross-validated *r*² results in fewer components and with less than 50% of the original variable set.

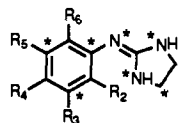
Table XII highlights many of the trends seen in the statistical results. The two- and three-Gaussian electrostatic evaluations of the Carbo index are seen to perform well in comparison to the corresponding grid-based calculations. This is highlighted by the results for series a(i), c, d, and e (see Table IX). The one-Gaussian evaluation is shown to be significantly inferior to the two- and three-Gaussian calculations (Table IX, series a(i) and d highlight this trend). The overall results for the Hodgkin and Carbo index evaluation of both electric field and electrostatic potential suggests there is little difference in

Table V. Structure and Activity Data for MPTP and MPTP Analogues of Series e^a

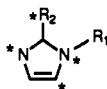


no.	structure	R ₁	R ₃	R ₄	R ₅	relative activity
1	A	Me	H	phenyl	H	3
2	A	H	H	phenyl	H	3
3	A	Me	Me	phenyl	H	2
4	A	Me	H	phenyl	Me	2
5	A	Me	H	2'-methylphenyl	H	3
6	A	Me	H	3'-methylphenyl	H	3
7	A	Me	H	4'-methylphenyl	H	2
8	A	Me	H	2'-methoxyphenyl	H	3
9	A	Me	H	3'-methoxyphenyl	H	2
10	A	Me	H	2'-flourophanyl	H	3
11	A	Me	H	3'-flourophanyl	H	3
12	A	Me	H	4'-flourophanyl	H	2
13	A	H	H	4'-flourophanyl	H	2
14	A	Me	H	2'-(trifluoromethyl)phenyl	H	3
15	A	Me	H	3'-chlorophenyl	H	3
16	A	Me	H	4'-chlorophenyl	H	2
17	A	H	H	4'-chlorophenyl	H	2
18	A	Me	H	cyclohexyl	H	3
19	A	Me	H	2'-thienyl	H	2
20	A	Me	H	benzyl	H	3
21	A	Me	H	tert-butyl	H	2
22	A	Me	H	ethyl carboxylate	H	2
23	A	Me	H	Et	H	2
24	A	Et	H	phenyl	H	2
25	A	2-hydroxy-Et	H	phenyl	H	2
26	A	isopropyl	H	phenyl	H	1
27	A	propyl	H	phenyl	H	1
28	A	butyl	H	phenyl	H	1
29	A	cyclopropylmethyl	H	phenyl	H	1
30	A	allyl	H	phenyl	H	1
31	A	benzyl	H	phenyl	H	1
32	A	Et	H	H	H	2
33	B					3
34	C	Me		2'-ethylphenyl		3
35	C	Me		2'-chlorophenyl		3
36	C	Me		3'-bromophenyl		3
37	C	Me		2',6'-dimethylphenyl		3
38	C	Me		1'-methylpyrrol-2'-yl		3
39	C	4'-cyanobutan-1'-one		phenyl		1
40	D	Me		phenyl		2
41	E	Me		phenyl		1

^a The atom of the R₄ group (or equivalent) α to the tetrahydropyridine ring was used in rigid fitting. Carbon atoms in any R₄ phenyl groups also used.

Table VI. Structural and pK_a Data for the Imidazolines of Series f

no.	R ₂	R ₃	R ₄	R ₅	R ₆	observed pK_a
1	Cl	Cl	H	H	H	8.55
2	Br	H	Br	H	Br	7.46
3	Cl	H	Cl	H	Cl	7.75
4	CH ₃	H	CH ₃	H	CH ₃	10.78
5	Cl	H	Cl	H	H	8.73
6	CH ₃	H	CH ₃	H	H	10.56
7	Cl	H	H	Cl	H	8.50
8	Br	H	H	H	Br	7.80
9	Cl	H	Br	H	Cl	7.72
10	Cl	H	CH ₃	H	Cl	8.29
11	Cl	H	NO ₂	H	Cl	6.86
12	Cl	H	OCH ₃	H	Cl	8.57
13	Cl	H	H	H	Cl	8.05
14	C ₂ H ₅	H	H	H	C ₂ H ₅	10.61
15	F	H	H	H	F	8.18
16	CH ₃	H	Br	H	CH ₃	10.21
17	CH ₃	H	Cl	H	CH ₃	10.25
18	CH ₃	H	H	H	CH ₃	10.53
19	Cl	H	CH ₃	H	H	9.41
20	Cl	H	H	H	H	9.15
21	CH ₃	H	Cl	H	H	9.99
22	CH ₃	H	H	H	H	10.23
23	H	H	H	H	H	10.05

Table VII. Structural and pK_a Data for the Imidazolines of Series g^a

no.	R ₁	R ₂	observed pK_a
1	CH ₃	Br	3.82
2	CH ₃	F	2.30
3	CH ₃	H	7.12
4	CH ₃	NH ₂	8.54
5	CH ₃	NO ₂	-0.48
6	H	Br	3.79
7	H	Cl	3.55
8	H	C ₂ H ₅	7.73
9	H	F	2.40
10	H	H	6.99
11	H	CH ₃	7.86
12	H	NH ₂	8.46
13	H	NO ₂	-0.81
14	H	C ₆ H ₅	6.48
15	H	NC ₆ H ₅	5.36
16	H	SCH ₃	5.95

^a Atom in R₂ group α to imidazole ring used in rigid fitting.

their general performance. The exponential and linear indices seem to provide similar performance and also appear superior to the Carbo and Hodgkin indices (see also Table IX series a and d). The Spearman rank correlation coefficient seems to work least well of all the evaluation techniques. Electric field evaluations performed well and appear to yield additional descriptive information for certain series (see Table IX series c and e). It would be of interest to extend the use of the linear and exponential indices to the evaluation of electric field.

For shape, the grid-based evaluations seem to perform slightly better than the corresponding Gaussian calculations. The major difference is seen for series a(i) of Table X. Nevertheless, Gaussian shape calculations perform well overall, with little difference seen in their individual behavior. Note that for average shape statistics the series

Table VIII. Substituents and Hammett Constant Data for Substituted Benzoic Acids of Series h^a

no.	substituent	Hammett constant	no.	substituent	Hammett constant
1	H	0.00	26	<i>p</i> -Br	0.23
2	<i>m</i> -Br	0.39	27	<i>p</i> -CF ₃	0.54
3	<i>m</i> -CF ₃	0.43	28	<i>p</i> -CH ₃	-0.17
4	<i>m</i> -CH ₃	-0.07	29	<i>p</i> -Cl	0.23
5	<i>m</i> -Cl	0.37	30	<i>p</i> -CN	0.66
6	<i>m</i> -CN	0.56	31	<i>p</i> -F	0.06
7	<i>m</i> -F	0.34	32	<i>p</i> -I	0.18
8	<i>m</i> -I	0.35	33	<i>p</i> -NH ₂	-0.66
9	<i>m</i> -NH ₂	-0.16	34	<i>p</i> -NO ₂	0.78
10	<i>m</i> -NO ₂	0.71	35	<i>p</i> -OCF ₃	0.35
11	<i>m</i> -OCF ₃	0.38	36	<i>p</i> -OH	-0.37
12	<i>m</i> -OH	0.12	37	<i>p</i> -OCH ₃	-0.27
13	<i>m</i> -OCH ₃	0.12	38	<i>p</i> -SH	0.15
14	<i>m</i> -SH	0.25	39	<i>p</i> -SCH ₃	0.00
15	<i>m</i> -SCH ₃	0.15	40	<i>p</i> -SCF ₃	0.50
16	<i>m</i> -SCF ₃	0.40	41	<i>p</i> -C(CH ₃) ₃	-0.20
17	<i>m</i> -C(CH ₃) ₃	-0.10	42	<i>p</i> -C ₂ F ₅	0.52
18	<i>m</i> -C ₂ F ₅	0.47	43	<i>p</i> -CH ₂ Br	0.14
19	<i>m</i> -CH ₂ Br	0.12	44	<i>p</i> -CH ₂ Cl	0.12
20	<i>m</i> -CH ₂ Cl	0.11	45	<i>p</i> -CH ₂ I	0.11
21	<i>m</i> -CH ₂ I	0.10	46	<i>p</i> -CH ₂ H ₅	-0.15
22	<i>m</i> -C ₂ H ₅	-0.07	47	<i>p</i> -SO ₂ CF ₃	0.93
23	<i>m</i> -SO ₂ CF ₃	0.79	48	<i>p</i> -SO ₂ F	0.91
24	<i>m</i> -SO ₂ F	0.80	49	<i>p</i> -SO ₂ CH ₃	0.72
25	<i>m</i> -SO ₂ CH ₃	0.60			

^a Acid group and benzene ring carbons used for rigid fitting.

d (cocaine analogue) data was ignored for reduced-set calculations. GOLPE showed the four strongest signal variables to be outside the reduced data set and was unable to remove any variables from the reduced set using progressive exclusion. The data shown in Table X were for the two variables GOLPE fixed as signal. These results were considered to be a poor model and were thus excluded from the averaging calculation. The results for variable grid increment evaluations were interesting. Previous studies have suggested somewhat conflicting values for the grid increments required for similarity calculations. For electrostatics, a 1.0-Å increment was said to be best together with a 4.0-Å extent.³ For shape, Meyer¹³ recommended an increment of 0.2 Å, while a different study⁹ suggested that an increment of 0.5 Å was sufficient. In this study the results suggest that an increment of 2.0 Å is sufficient for electronic calculations, with results virtually identical to those achieved with a 1.0-Å increment. While the average result for the 3.0-Å grid is good, Table XI results illustrate a number of problems with this increment. Series a(i) and d show significantly reduced performance, while for series e the program would not run with a 3.0-Å increment, as no grid points with an electrostatic potential magnitude of 5 kcal mol⁻¹ or greater could be located. Additional tests using a 2.0-Å grid extent together with a 2.0-Å increment were also run. The results obtained were virtually identical to those produced with a 3.0-Å increment and 4.0-Å extent. The minimum required configuration would thus appear to be a 2.0-Å/4.0-Å extent for *N* by *N* single point calculations. It is probable that denser increments will be required for similarity optimizations to lower the risk of premature convergence, since numerical evaluations of Carbo and Hodgkin index integrals will be very rough with a 2.0-Å/4.0-Å system. It should be noted, however, that, as in previous investigations,⁷ these studies point strongly to the use of the analytical Gaussian evaluations as the one of choice for similarity optimizations. For shape, Tables XI and XII show that the 0.2- and 0.5-Å results are virtually

Table IX. Statistical Data for Electrostatic Potential and Electric Field Similarity Calculations^a

evaluation method	statistical data					
	series a(i) steroids and TBG data		series a(ii) steroids and CBG data		series b triazines	
Carbo	0.25/2	0.38/2	0.49/2	0.56/2	0.08/4	0.27/3
3-Gaussian Carbo	0.75/5	0.79/3	0.53/2	0.59/1	0.19/3	0.24/2
2-Gaussian Carbo	0.75/7	0.76/3	0.52/2	0.58/1	0.16/3	0.23/2
1-Gaussian Carbo	0.34/2	0.40/2	0.49/2	0.55/2	0.04/3	0.13/2
Hodgkin	0.51/4	0.47/3	0.53/2	0.62/2	0.10/4	0.34/2
exponential	0.61/4	0.68/3	0.74/2	0.73/1	0.21/2	0.30/2
linear	0.59/4	0.65/3	0.66/1	0.74/1	0.15/3	0.31/3
Spearman	0.04/2	0.20/2	0.34/2	0.39/1	0.04/1	0.1/1
Carbo field	0.64/5	0.77/3	0.62/1	0.69/1	0.47/7	0.39/3
Hodgkin field	0.64/5	0.72/3	0.63/1	0.69/1	0.47/7	0.46/3
	series c β -carbolines, etc.		series d cocaine analogues		series e MPTP and analogues	
Carbo	0.41/5	0.42/3	0.37/1	0.38/1	no predictivity	0.13/3
3-Gaussian Carbo	0.56/5	0.56/3	0.58/5	0.57/3	0.30/2	0.52/3
2-Gaussian Carbo	0.56/5	0.57/3	0.55/5	0.59/3	0.25/2	0.42/2
1-Gaussian Carbo	0.55/5	0.46/3	0.27/1	0.31/1	0.16/4	0.34/2
Hodgkin	0.59/5	0.56/3	0.37/1	0.35/1	no predictivity	0.14/2
exponential	0.41/3	0.48/3	0.46/1	0.51/1	0.01/3	0.26/3
linear	0.39/3	0.45/2	0.43/1	0.43/1	no predictivity	0.26/3
Spearman	0.22/5	0.26/3	0.38/1	0.46/1	no predictivity	0.16/2
Carbo field	0.53/3	0.61/3	0.25/1	0.42/1	0.21/3	0.47/3
Hodgkin field	0.62/3	0.64/3	0.27/1	0.45/1	0.20/3	0.40/3
	series f imidazolines		series g imidazoles		series h benzoic acids	
Carbo	0.96/8	0.91/3	0.56/2	0.79/2	0.80/4	0.70/1
3-Gaussian Carbo	0.86/4	0.93/3	0.85/2	0.91/2	0.83/2	0.85/2
2-Gaussian Carbo	0.94/4	0.94/3	0.86/3	0.88/3	0.80/5	0.81/3
1-Gaussian Carbo	0.83/3	0.91/2	0.88/3	0.89/3	0.88/8	0.81/3
Hodgkin	0.91/4	0.89/3	0.63/3	0.77/3	0.86/7	0.70/1
exponential	0.62/2	0.89/3	0.64/3	0.82/2	0.83/7	0.82/3
linear	0.60/3	0.82/3	0.67/3	0.87/2	0.81/4	0.85/3
Spearman	0.48/2	0.90/3	0.71/3	0.86/1	0.77/4	0.65/3
Carbo field	0.87/2	0.90/2	0.76/2	0.84/2	0.87/4	0.89/3
Hodgkin field	0.87/3	0.90/2	0.80/2	0.83/2	0.87/4	0.88/3

^a For each series two sets of results are given. The first set shows the cross-validated r^2 and the associated number of PLS components for the full matrix. The second set shows the cross-validated r^2 and the associated number of PLS components for the matrix after variable selection with GOLPE.

Table X. Statistical Data for Shape Similarity Calculations^a

evaluation method	statistical data					
	series a(i) steroids + TBG data		series a(ii) steroids + CBG data		series b triazines	
unmodified Gaussians	0.12/5	0.14/3	0.71/5	0.81/3	0.61/5	0.65/2
H-VDW fitted Gaussians	0.34/4	0.37/3	0.72/3	0.77/3	0.62/4	0.76/3
VDW fitted Gaussians	0.34/4	0.37/3	0.72/3	0.79/3	0.64/3	0.70/3
grid-based reduced set	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	0.76/6	0.74/3
grid-based full set	0.4/4	0.59/3	0.76/3	0.82/3	0.68/7	0.68/3
	series c β -carbolines, etc.		series d cocaine analogues		series e MPTP + analogues	
unmodified Gaussians	0.45/3	0.55/3	0.10/1	0.50/1	0.23/2	0.49/3
H-VDW fitted Gaussians	0.49/3	0.58/3	no predictivity	0.14/1	0.22/2	0.55/3
VDW fitted Gaussians	0.51/3	0.58/1	no predictivity	no predictivity	0.23/4	0.51/3
grid-based reduced set	0.52/6	0.60/2	0.20/6	0.30/1	0.27/1	0.52/3
grid-based full set	0.60/3	0.65/3	no predictivity	0.71/3	0.38/2	0.52/2

^a For each series two sets of results are given. The first set shows the cross-validated r^2 and the associated number of PLS components for the full matrix. The second set shows the cross-validated r^2 and the associated number of PLS components for the matrix after variable selection with GOLPE. ^b Reduced and full set identical for steroids.

identical. 1.0- and 2.0-Å increments show similar overall performance, but results for series a(i) are inferior. Series d shows an apparent anomaly in that the 1.0-Å increment results are inferior to those of the 2.0-Å increment. A study of the signal variables shows that those chosen for the 2.0-Å increment are significantly different from the other systems, suggesting a chance correlation of some form. Overall it would seem that a 0.5-Å increment is sufficient to guarantee similar results to those of a denser

grid. Coarser grids perform fairly well but can behave less reliably.

The results obtained for the mixed matrices (Table XIII) compare well with those achieved using CoMFA. It is unfair to make direct comparisons between results since different statistical packages and procedures have been used to derive the data. Nevertheless it is encouraging that the similarity models do so well. It should be noted that on average fewer than 10 variables remained in the

Table XI. Statistical Data using Hodgkin Electric Field and Meyer Type Shape Similarity Calculations with Different Grid Increments^a

evaluation method and increment	statistical data					
	series a(i) steroids and TBG data		series a(ii) steroids and CBG data		series b triazines	
Hodgkin field, 1.0 Å	0.64/5	0.72/3	0.63/1	0.69/1	0.47/7	0.46/3
Hodgkin field, 2.0 Å	0.59/3	0.70/3	0.59/1	0.65/1	0.41/7	0.4/3
Hodgkin field, 3.0 Å	0.40/2	0.51/2	0.57/2	0.65/1	0.27/6	0.4/3
grid-based full set, 0.2 Å	0.38/4	0.58/3	0.77/3	0.82/3	0.69/7	0.70/3
grid-based full set, 0.5 Å	0.40/4	0.59/3	0.76/3	0.82/3	0.68/7	0.68/3
grid-based full set, 1.0 Å	0.20/6	0.35/3	0.75/3	0.81/3	0.64/6	0.67/3
grid-based full set, 2.0 Å	no predictivity	0.32/3	0.77/3	0.80/3	0.60/4	0.63/3
	series c β-carbolines, etc.		series d cocaine analogues		series e MPTP and analogues	
Hodgkin field, 1.0 Å	0.62/3	0.64/3	0.27/3	0.45/1	0.20/3	0.40/1
Hodgkin field, 2.0 Å	0.66/5	0.63/3	0.20/1	0.52/2	no predictivity	0.36/3
Hodgkin field, 3.0 Å	0.43/5	0.53/3	0.20/1	0.35/3	<i>b</i>	<i>b</i>
grid-based full set, 0.2 Å	0.60/3	0.66/3	no predictivity	0.72/3	0.37/2	0.53/2
grid-based full set, 0.5 Å	0.60/3	0.65/3	no predictivity	0.71/3	0.38/2	0.52/2
grid-based full set, 1.0 Å	0.60/3	0.65/3	no predictivity	0.41/3	0.38/2	0.52/2
grid-based full set, 2.0 Å	0.52/3	0.61/3	0.58/6	0.67/3	0.44/2	0.58/2
	series f imidazolines		series g imidazoles		series h benzoic acids	
Hodgkin field, 1.0 Å	0.87/2	0.90/2	0.80/2	0.83/3	0.87/4	0.88/2
Hodgkin field, 2.0 Å	0.88/4	0.89/3	0.72/2	0.84/2	0.85/7	0.85/3
Hodgkin field, 3.0 Å	0.92/3	0.93/3	0.71/2	0.84/3	0.66/3	0.70/3

^a For each series two sets of results are given. The first set shows the cross-validated r^2 and the associated number of PLS components for the full matrix. The second set shows the cross-validated r^2 and the associated number of PLS components for the matrix after variable selection with GOLPE. ^b Program would not run for MPTP with 3.0-Å grid increment, as no grid point could be found with an electrostatic potential magnitude of 5 kcal mol⁻¹ or greater.

Table XII. Average Cross-Validated r^2 Values for the Various Index Evaluations

evaluation method	average cross-validated r^2	evaluation method	average cross-validated r^2
Carbo	0.50	Hodgkin field, 2.0 Å	0.65
3-Gaussian Carbo	0.66	Hodgkin field, 3.0 Å	0.61
2-Gaussian Carbo	0.64	unmodified Gaussians ^a	0.60
1-Gaussian Carbo	0.54	H VDW-fitted Gaussians ^a	0.61
Hodgkin exponential	0.53	VDW-fitted Gaussians ^a	0.59
linear	0.61	grid-based reduced set ^a	0.65
Spearman	0.45	grid-based full set	0.66
Carbo field	0.66	grid-based full set, 0.2 Å	0.67
Hodgkin field	0.67	grid-based full set, 1.0 Å	0.57
		grid-based full set, 2.0 Å	0.60

^a Results for series d (cocaine analogues) excluded from the average calculation. See Discussion.

final mixed matrix model for each series. This is the main advantage of similarity over CoMFA which often contains hundreds of variables in its statistical models. The small size of the data matrices means that PLS model cross-validations and optimizations can be swiftly accomplished. This is highlighted by the fact that all the progressive exclusion calculations took less than 5 min to complete on a Silicon Graphics R4000 Indigo, even with the rigorous design ratios set. When this speed is combined with that of the Gaussian similarity evaluations, QSAR model generation becomes very rapid indeed. The primary disadvantage of the similarity models is that no automatic graphical analysis of the resulting model is possible. Interpretation of the model is therefore more complicated.

Conclusions

The results of these studies suggest that similarity matrices can be used to derive good QSAR models for many different systems. Analytical Gaussian evaluations are shown to provide good biological descriptive power (especially for electrostatics). Electric field calculations appear to provide additional information for some systems.

Table XIII. Statistical Data for Combined Electronic and Shape Similarity Data Matrices^a

series	unmodified matrix	GOLPE-reduced matrix	CoMFA results
a(i), steroids and TBG data	0.73/3	0.77/2	0.44/4
a(ii), steroids and CBG data	0.79/2	0.82/2	0.69/2
b, triazines	0.74/5	0.73/3	0.47/2
c, β-carbolines, etc.	0.69/4	0.72/3	0.59/4
d, cocaine analogues	0.54/2	0.64/2	0.57/4
e, MPTP and analogues	0.50/2	0.56/2	0.57/4
f, imidazolines ^b		0.32	0.27
g, imidazoles ^b		0.90	0.69
h, benzoic acids ^b		0.12	0.05

^a For each series three sets of results are given. The first set shows the cross-validated r^2 and the associated number of PLS components for the full combined matrix. The second set shows the cross-validated r^2 and the associated number of PLS components for the matrix after variable selection with GOLPE. The final set shows the equivalent statistical data obtained by CoMFA studies of the same series. ^b Only electronic matrix data used for these series. Cross-validated r^2 data not provided in original CoMFA papers, so predictive standard error listed instead.

Grid-based shape evaluations and electrostatic evaluations using the linear and exponential indices also work well. Grid increments of 0.5 Å for shape and 2.0 Å for electrostatics would appear sufficient to describe the biological systems studied.

It is hoped that the studies undertaken here will stimulate further work into the use of similarity data for the elucidation of QSAR models.

Acknowledgment. This work is supported through an SERC CASE award with British Biotechnology Ltd, Oxford, England.

References

- (1) Hopfinger, A. J. A QSAR Investigation of DHFR Inhibition by Bakers Triazines Based Upon Molecular Shape Analysis. *J. Am. Chem. Soc.* 1980, 102, 7196-7206.

- (2) Hopfinger, A. J. Theory and Analysis of Molecular Potential Energy Fields in Molecular Shape Analysis: A QSAR Study of 2,4-diamino-5-benzylpyrimidines as DHFR Inhibitors. *J. Med. Chem.* 1983, 26, 990-996.
- (3) Burt, C.; Huxley, P.; Richards, W. G. The Application of Molecular Similarity Calculations. *J. Comput. Chem.* 1990, 11, 1139-1146.
- (4) Good, A. C.; So, S.; Richards, W. G. Structure Activity Relationships from Similarity Matrices. *J. Med. Chem.* In press.
- (5) Cramer, R. D. III.; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Cross-validation, Bootstrapping, and Partial Least Squares Compared with Multiple Linear Regression in Conventional QSAR Studies. *Quant. Struct. Act. Relat.* 1988, 7, 18-25.
- (6) Cramer, R. D. III.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* 1988, 110, 5959-5967.
- (7) Good, A. C.; Hodgkin, E. E.; Richards, W. G. The Utilization of Gaussian Functions for the Rapid Evaluation of Molecular Similarity. *J. Chem. Inf. Comput. Sci.* 1992, 32, 188-191.
- (8) Good, A. C.; Richards, W. G. Rapid Evaluation of Shape Similarity Using Gaussian Functions. *J. Chem. Inf. Comput. Sci.* In press.
- (9) Burt, C.; Steric Alignment and Similarity Calculations in Terms of Molecular Shape in A Theoretical Investigation of Bioisosterism by Molecular Similarity, Ph.D. Thesis, Oxford University, Chapter 7.
- (10) Carbo, R.; Leyda, L.; Arnau, M. An Electron Density Measure of the Similarity Between Two Compounds. *Int. J. Quantum Chem.* 1980, 17, 1185-1189.
- (11) Carbo, R.; Domingo, L. LCAO-MO Similarity Measures and Taxonomy. *Int. J. Quantum Chem.* 1987, 32, 517-545.
- (12) Oxford Molecular Ltd, The Magdalen Centre, Oxford Science Park, Sandford on Thames, Oxford OX4 4GA, United Kingdom.
- (13) Meyer, A. M.; Richards, W. G. Similarity of Molecular Shape. *J. Comput.-Aided Mol. Design*, 1991, 5, 426-439.
- (14) Gaussian 88. Frisch, M. J.; Head, Gordon M.; Schlegel, H. B.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Martin, R.; Kahn, L. R.; Stewart, J. J. P.; Fluder, E. M.; Topiol, S.; Pople, J. A.; Gaussian, Inc., Pittsburgh, PA, 1988.
- (15) Hodgkin, E. E.; Richards, W. G. Molecular Similarity Based on Electrostatic Potential and Electric Field. *Int. J. Quantum Chem. Quantum Biol. Symp.* 1987, 14, 105-110.
- (16) Good, A. C. The calculation of Molecular Similarity: alternative Formulas, data manipulation and graphical display. *J. Mol. Graph.* 1992, 10, 144-151.
- (17) Manaut, M.; Sanz, F.; Jose, J.; Milesi, M. Automatic Search for Maximum Similarity Between Molecular Electrostatic Potential Distributions. *J. Comput.-Aided Mol. Des.* 1991, 5, 371-380.
- (18) McFarland, J. W. Comparative Molecular Field Analysis of Anticoccidial Triazines. *J. Med. Chem.* 1992, 35, 2543-2550.
- (19) Allen, M. S.; Tan, Y.; Trudell, M. L.; Narayanan, K.; Schindler, L. R.; Martin, M. J.; Scultz, C.; Hagen, T. J.; Koehler, K. F.; Coddling, P. W.; Skolnick, P.; Cook, J. M. Synthetic and Computer Assisted Analysis of the Pharmacophore for the Benzodiazepin Receptor Inverse Agonist Site. *J. Med. Chem.* 1990, 33, 2343-2357.
- (20) Carroll, F. I.; Gao, Y.; Rahman, M. A.; Abraham, P.; Parham, K.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J. Synthesis, Ligand Binding, QSAR and CoMFA Study of 3 β -(p-substituted phenyl)tropane-2 β carboxylic acid methyl esters. *J. Med. Chem.* 1991, 34, 2719-2725.
- (21) Maret, G. M.; Tayar, N. E.; Carrupt, P. A.; Testa, B.; Jenner, P.; Baird, M. Toxication of MPTP and Analogues by Monoamine Oxidase. *Biochem. Pharm.* 1990, 40, 783-792.
- (22) Kim, K. H.; Martin, Y. C. Direct Prediction of Dissociation Constants of Clonidine-like Imidazolines, 2-Substituted Imidazoles and 1-Methyl-2-substituted-imidazoles from 3D Structures Using a Comparative Molecular Field Analysis Approach. *J. Med. Chem.* 1991, 34, 2056-2060.
- (23) Kim, K. H.; Martin, Y. C. Direct Prediction of Linear Free Energy Substituent Effects from 3D Structures Using Comparative Molecular Field Analysis. 1. Electronic Effects of Substituted Benzoic Acids. *J. Org. Chem.* 1991, 56, 2723-2729.
- (24) Allen, F. N.; Kennard, O.; Taylor, R. Systematic Analysis of Structural Data as a Research Technique in Organic Chemistry. *Acc. Chem. Res.* 1983, 146-153.
- (25) CHEMX, Chemical Design Ltd, Roundway House, Cromwell Buisness Park, Chipping Norton, Oxon, OX7 5SR.
- (26) MOPAC, Stewart, J. J. P. *QPCE* 455.
- (27) Ferenzy, G.; Reynolds, C. A.; Richards, W. G. Semi-empirical AM1 Electrostatic Potential and AM1 Electrostatic Potential Derived Charges, a Comparison with Ab initio Values. *J. Comput. Chem.* 1990, 11, 159-159.
- (28) Baroni, M.; Constantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. GOLPE: an Advanced Chemometric Tool for 3D-QSAR Problems. *Quant. Struct. Act. Relat.* 1993, 12, 9-20.
- (29) Mitchell, T. J. An Algorithm for the Construction of D-Optimal Experimental Designs. *Technometrics* 1974, 16, 203-210.
- (30) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters*; Wiley J. & Sons: New York, 1978; Chapter 12.