

Articles

Comparative Molecular Moment Analysis (CoMMA): 3D-QSAR without Molecular Superposition

B. D. Silverman* and Daniel. E. Platt

IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

Received August 7, 1995[§]

3d-QSAR procedures utilize descriptors that characterize molecular shape and charge distributions responsible for the steric and electrostatic nonbonding interactions intimately involved in ligand–receptor binding. Comparative molecular moment analysis (CoMMA) utilizes moments of the molecular mass and charge distributions up to and including second order in the development of molecular similarity descriptors. As a consequence, two Cartesian reference frames are then defined with respect to each molecular structure. One frame is the principal inertial axes calculated with respect to the center-of-mass. For neutrally charged molecular species, the other reference frame is the principal quadrupolar axes calculated with respect to the molecular “center-of-dipole”. QSAR descriptors include quantities that characterize shape and charge independently as well as quantities that characterize their relationship. 3D-QSAR partial least squares (PLS) cross-validation procedures are utilized to predict the activity of several training sets of molecules previously investigated. This is the first time that molecular electrostatic quadrupolar moments have been utilized in a 3D-QSAR analysis, and it is shown that descriptors involving the quadrupolar moments and related quantities are required for the significant cross-validated predictive r^2 s obtained. CoMMA requires no superposition step, i.e., no step requiring a comparison between two molecules at any stage of the 3D-QSAR calculation.

1. Introduction

Fundamental to the implementation of any three dimensional quantitative structure–activity relationship (3D-QSAR) study is an ability to characterize the shape and charge distributions of a molecule in three-dimensional space. This is a consequence of the predominance of steric and electrostatic interactions in the binding of a drug to its targeted receptor site. One expects that molecules with shape and charge distributions similar to a molecule or set of molecules exhibiting a desired biological response will be optimal candidates for further study in connection with such response.

Molecular shape and charge distributions with their consequent steric and electrostatic profiles have been characterized a number of different ways. Molecular shape may be assigned by simply superimposing van der Waals spheres about the respective atoms that compose the molecule. Steric energies may be calculated from Lenard-Jones 6–12 or other phenomenological potentials. Molecular shape may also be determined from an examination of the electronic density distribution obtained from a molecular orbital or other such quantum chemistry calculation. Electrostatic energies have been assigned by calculating Coulomb sums utilizing point charges either at or in the vicinity of molecular atomic sites or implicitly assigned by identifying molecular regions where significant electrostatic energies are predominant, such as potential hydrogen bond donor and acceptor locations. Such characterizations of molecular shape and charge involve a detailed description of local spatial variation or position and consequently

exhibit no succinct relationship to either the underlying molecular structure or each other. As a consequence, steric and electrostatic molecular similarity can only be assigned after molecules are superimposed and registered in some manner which minimizes differences in the distribution being examined. There are methods utilizing atom-pair distances which do not require molecular superposition, for example, GEOSIM.¹ Such methods have, however, been predominantly utilized for large-scale screening.

A simple and fundamental characterization of molecular shape and charge is given by the moments of the shape and charge distributions. At present we will characterize the shape of a molecule by a characterization of its mass. The lower order moments are clearly understood. The zeroth-order moments of the mass and charge distributions are nothing more than the total molecular mass and net molecular charge. The first-order moment of the mass distribution is a quantity not usually mentioned, since if the origin of the coordinate system is chosen at the center of mass, this quantity is zero. In fact the center-of-mass is so defined to zero out the first-order moment of the mass distribution. The second-order moments of the mass distribution are the moments of inertia. These moments together with the principal inertial axes provide important information to be utilized in developing molecular descriptors. Section 2 will review some relevant material.

The first-order moment of the charge distribution is the dipole moment. For neutral molecules, this is unchanged or invariant with respect to the location of the origin of the axes about which it is calculated. This invariance is a consequence of the fact that the lowest

[§] Abstract published in *Advance ACS Abstracts*, May 1, 1996.

order nonvanishing moment of the electrostatic multipolar expansion does not depend upon the origin about which it is calculated. The values of all higher order multipolar moments do depend upon the choice of origin of the multipolar expansion. Therefore, if one was inclined to utilize moments of the mass and charge distributions for neutral molecules up to and including second order, one would be faced with a problem. How should the second-order moments of the charge distribution be calculated, namely, the electrostatic quadrupolar moments, since they are dependent upon the choice of origin of the axes about which they are calculated?

Prior to a recent investigation,² no procedure had identified an origin of electrostatic multipolar expansion for molecules of arbitrary symmetry and structure that could be utilized for the purpose of molecular similarity comparison between the moments of order higher than the lowest order nonvanishing moment. For lowest order nonvanishing moment of order N , a "center-of- N -pole" was identified and utilized for molecular similarity comparison between molecules with identical order of lowest nonvanishing multipolar moment. It was emphasized that, for molecules with zero net charge and nonvanishing dipole moment, quadrupolar principal values and axes can only be utilized for similarity comparison if the multipolar expansion is performed about the appropriate expansion center, namely, the "center-of-dipole." Section 3 will review material involving the center-of-dipole relevant for the present study of neutrally charged molecules with nonvanishing dipole moment. Generalization to center-of- N -pole appropriate for ionic as well as neutral nonpolar molecules can be found in the earlier reference.² This procedure enables one to define molecular similarity unambiguously to arbitrary order of the electrostatic multipolar expansion.

Section 4 will enumerate the set of molecular moment descriptors utilized in the 3D-QSAR investigations to be presented in the present paper, and Section 5 will outline the computational procedures. Section 6 will examine the results of applying this procedure to the prediction of the biological and/or chemical reactivity of several sets of molecules previously investigated. Sections 7 and 8 will be devoted to discussion and conclusions.

Sections 2 and 3 have been included for completeness. They may be either skipped or read cursorily before proceeding to Section 4.

2. Center of Mass/Gravity: Inertial Moments

The angular momenta of a rigid body about any point in space can be written as a sum of two contributions, one arising from the translational motion of the center of mass and the other from the rotational motion about this center.³ The center-of-mass provides a unique reference origin that enables one to simply separate the translational motion of a rigid body or molecule from its rotational motion. The angular momentum \vec{L} about the center of mass is linearly related to the components of angular velocity $\vec{\omega}$ by the three-dimensional second-rank moment of inertia tensor \mathbf{I}

$$\vec{L} = \mathbf{I} \cdot \vec{\omega}$$

This tensor involves second-order spatial moments of the mass distribution of the molecule. The components

of the inertial tensor can be written

$$\mathbf{I} = \sum_i m_i (\mathbf{1} r_i^2 - \vec{r}_i \vec{r}_i)$$

where i is summed over the atomic centers, m_i is the atomic weight of the i th atom, \vec{r}_i is a vector from the center of rotation to the i th atom.

There are three special directions, or principal axes, for which the angular velocities about these directions are parallel to their respective components of angular momenta. If the moment of inertia tensor is represented in terms of these principal axes, the tensor will be diagonal. The diagonal moments corresponding to these principal axes are called the principal inertial moments. Since these directions and principal inertial moments involve a characterization of molecular orientation and shape, we will use these quantities as molecular shape descriptors. These moments depend, however, on the center of expansion. Even for identical molecules, the principal values of the moments, as well as their principal axes, depend upon the center of expansion. Since the origin is otherwise arbitrary, some way of selecting a unique center that carries the same information for all molecules must be defined. Once such a center has been defined, the moments may then be used for a similarity comparison of molecular features measured by the moments. The moments of the mass distribution are calculated about the unique center-of-mass of each molecule.

It should be noted that the molecular second-order spatial moments composing the inertial tensor are weighted by the atomic mass at each of the atomic sites. One might question whether this weighting would be as representative of overall molecular shape as would spatial moments weighted either equally at the atomic centers or perhaps in some other manner to mirror overall molecular shape. Using the inertial tensor does provide 3D-QSAR descriptors which are directly related to certain kinematical and dynamical aspects of molecular binding. The ultimate test of how such moments should be weighted will depend upon the effect that such weighting has upon the calculated predictability of molecular biological or chemical activity.

3. Center of Dipole: Quadrupolar Moments

The charge distribution of a molecule can be characterized by its multipolar components.⁴ These components provide a description of the far-field electrostatic potential at position \vec{x} of an arbitrary charge distribution $\rho(\vec{x})$ since they appear in an expansion of the electrostatic potential $V(\vec{x})$ in powers of a small parameter. The expansion follows from

$$V(\vec{x}) = \int d^3x' \frac{\rho(\vec{x}')}{|\vec{x} - \vec{x}'|}$$

and

$$\frac{1}{|\vec{x} - \vec{x}'|} = \sum_{lm} \frac{r_{<}^l}{r_{>}^{l+1}} Y_{lm}'(\hat{x}') Y_{lm}(\hat{x})$$

where $Y_{lm}(\hat{x})$ is the spherical harmonic indexed with rotational symmetry indices l and m in the direction $\hat{x} = \vec{x}/|\vec{x}|$. $r_{<} = \min\{|\vec{x}|, |\vec{x}'|\}$, and $r_{>} = \max\{|\vec{x}|, |\vec{x}'|\}$. The

expansion parameter, $r_</r_>$, is essentially the ratio of the distance of the charge distribution from the origin of expansion (or extent of the charge distribution) to the distance between the far field point and origin of expansion. The far-field electrostatic potential written in terms of the multipolar components is

$$V(\vec{x}) = \sum_{lm} \frac{q_{lm}}{r^{l+1}} Y_{lm}(\hat{x})$$

where

$$q_{lm} = \int d^3x \rho(\vec{x}) r^l Y_{lm}^*(\hat{x})$$

q_{00} refers to the total monopole charge. For neutral molecules, this is zero. q_{1m} for $m = -1, 0, 1$ refers to the dipole contribution. For nonpolar molecules, this is zero. The Cartesian form of this is the vector

$$\vec{p} = \int d^3x \vec{x} \rho(\vec{x}).$$

For $l = 2$, the q_{lm} refer to the quadrupolar contributions. The increase in angular dependency with increasing index, l , of each term in the potential energy expansion is accompanied by a more rapid drop off with increasing r in the far-field region. The spherical quadrupolar components q_{2m} 's may be expressed more effectively for the purposes of extracting principal values and axes in Cartesian form as

$$\mathbf{Q} = \frac{1}{2} \int d^3x (3\vec{x}\vec{x} - |\vec{x}|^2 \mathbf{1}) \rho(\vec{x})$$

The quadrupolar moments/components compose a three-dimensional second rank tensor which in diagonal form provides a small set of numbers, namely, the principal quadrupolar components plus the orientation of the principal quadrupolar axes with respect to the underlying molecular structure. One might, therefore, expect these to be utilized as molecular charge descriptors in a fashion analogous to the inertial moments described in the previous section. The difficulty, however, is that the only translationally invariant multipole moment of the multipolar expansion is the lowest order nonvanishing moment. Consequently, for a neutral molecule, the only translationally invariant moment is the dipole moment. The components of the quadrupolar tensor will then depend upon the origin of multipolar expansion. A "magic-center" or "center-of-dipole" can, however, be defined² for neutrally charged molecules that will guarantee that two identical charge distributions, arbitrarily displaced and rotated in space with respect to each other, will yield multipolar moments that are identical up to and beyond quadrupolar order if the multipolar expansions are performed about such center for each of the molecules. Such center provides, therefore, a reference origin for the comparison of multipolar moments of an arbitrary set of molecules with vanishing net charge and nonvanishing dipole moment. This center, called the center-of-dipole, is obtained by finding the center of multipolar expansion which minimizes the solid angle average of the squared deviation of the total far-field potential from the dipolar contribution to this potential

$$\oint d\Omega(\hat{x}) \left| V(\vec{x}) - \sum_m \frac{q_{1m}}{r^2} Y_{1m}(\hat{x}) \right|^2$$

It places the origin of multipolar expansion at a point in space where the electrostatic dipolar field, in an averaged sense, most closely approximates the far field potential.

For a dipole moment vector, \vec{p} , and a quadrupolar tensor, \mathbf{Q} , calculated about an arbitrary origin, the displacement from this origin to the center-of-dipole is given by

$$\vec{d} = \frac{2}{3p^2} \left[\mathbf{Q} \cdot \vec{p} - \left(\frac{\vec{p} \cdot \mathbf{Q} \cdot \vec{p}}{4p^2} \right) \vec{p} \right]$$

At this center-of-dipole, one finds the dipole moment vector to be an eigenvector with zero eigenvalue of the quadrupolar operator.

$$\mathbf{Q}' \cdot \vec{p} = \vec{0},$$

where \mathbf{Q}' is the quadrupole moment tensor at the center of dipole. This yields an interesting invariant relationship between the dipolar direction in space and the principal quadrupolar axes about the center-of-dipole. This is illustrated in Figure 1. The dipole points along the quadrupolar principal axis that is associated with zero quadrupolar moment, and the two nonvanishing quadrupolar moments of equal magnitude and opposite sign are, therefore, along directions perpendicular to this direction.

Multipolar expansion about the center-of-dipole and rotation to principal axes orientation about this center, therefore, provides a few numbers, namely, dipolar and quadrupolar magnitudes as well as the quadrupolar principal axes orientation with respect to the underlying molecular structure that can be utilized as molecular electrostatic moment descriptors.

A center-of-dipole had been alluded to previously⁵ as the center at which the quadrupolar components of a linear molecule were zero. For arbitrary molecular symmetry, it is not possible to zero out the five independent quadrupolar components by translating the origin of expansion. The present procedure minimizes the quadrupolar contribution to the solid angle average of the square of the potential.

4. Molecular Moment Descriptors for 3D-QSAR

Moments of the molecular mass and charge distributions can, therefore, be calculated up to and including second order for the purpose of molecular similarity comparison. How might such moments be used to develop descriptors for a 3D-QSAR analysis? Thirteen such descriptors will be defined and utilized in the present paper. These descriptors may be classified in three different categories: descriptors that relate solely to molecular shape, descriptors that relate solely to molecular charge, and finally descriptors that involve the relationship between shape and charge.

Descriptors relating solely to molecular shape are the three principal moments of inertia, I_x , I_y , I_z . The two descriptors that relate solely to molecular charge are the magnitude of the dipole moment, p , and the magnitude of the principal quadrupole moment, Q . Descriptors that relate shape and charge can be developed in a

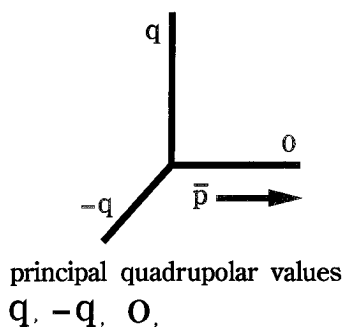


Figure 1. Principal quadrupolar axes and values about the center-of-dipole.

number of different ways. In the present work they have been chosen simply as follows: the magnitudes of the dipolar components as well as the magnitudes of the components of displacement between the center-of-mass and center-of-dipole are calculated with respect to the principal inertial axes. This provides six additional descriptors, namely p_x , p_y , p_z and d_x , d_y , d_z . Quadrupolar components are calculated with respect to a translated inertial reference frame whose origin coincides with the center-of-dipole. This provides two additional descriptors, Q_{xx} and Q_{yy} , that are invariant with respect to the sense of the inertial axes to which they are referenced. The tracelessness of the quadrupolar tensor precludes use of one of the diagonal tensor components as an independent variable.

This set of 13 numbers, therefore, provides a set of three-dimensional internal molecular moment descriptors that is invariant with respect to molecular rotation and translation in space, i.e., it is independent of the orientation and location of the molecules in three-dimensional space. Molecular superposition is therefore precluded from similarity comparisons between the descriptors of different molecules.

To complete the set of descriptors up to and including the second-order spatial moments of molecular mass, one might include the zeroth-order moment, namely, molecular mass. Molecular mass has not been utilized in the calculations presented in the following section. It might, however, be a useful descriptor in similarity searches addressing sets of molecules that exhibit a greater diversity of size than treated in the following sections. One might also utilize the magnitude of the displacement between the center-of-dipole and center-of-mass as well as its components. The magnitude is dependent on its components in a nonlinear manner, and as such contains linearly independent and distinct information. Finally, the two sets of principal axes, two vectors, and several scalars provide the opportunity of developing other descriptors involving the relationship of molecular charge to shape. More extensive combinations of descriptors have yet to be evaluated. In the present paper we limit ourselves to investigating the consequences of utilizing the minimal set of 13 descriptors described previously.

Dipole moments and their relationship to molecular structure as well as inertial components and axes have been utilized previously in connection with QSAR studies.^{6,7} As far as we are aware, the present paper is the first to propose simultaneous utilization of the zeroth first-, and second-order spatial moments of the charge as well as the mass distribution in connection with a QSAR analysis.

Table 1. Steroid Binding Affinity

entry no.	molecule	cbg (log(1/K))	tbg (log(1/K))
1	aldosterone	6.279	5.322
2	deoxycorticosterone	7.653	7.380
3	deoxycortisol	7.881	7.204
4	dihydrotestosterone	5.919	9.740
5	estradiol	5.000	8.833
6	estriol	5.000	6.633
7	estrone	5.000	8.176
8	etiocholanolone	5.255	6.146
9	pregnenolone	5.255	7.146
10	17-hydroxypregnenolone	5.000	6.362
11	progesterone	7.380	6.944
12	androstenediol	5.000	9.114
13	17-hydroxyprogesterone	7.740	6.996
14	testosterone	6.724	9.204
15	androstenediol	5.000	9.176
16	androstenedione	5.763	7.462
17	androsterone	5.613	7.146
18	corticosterone	7.881	6.342
19	cortisol	7.881	6.204
20	cortisone	6.892	6.431
21	dehydroepiandrosterone	5.000	7.819
22	1	7.512	
23	2	7.553	
24	3	6.779	
25	4	7.200	
26	5	6.144	
27	6	6.247	
28	7	7.120	
29	8	6.817	
30	9	7.688	
31	10	5.797	

5. Computational Methods

To test the ability of the set of the 13 internal CoMMA descriptors to predict molecular activity, the following five series have been investigated, four of which have been previously investigated by other 3D-QSAR procedures: (a) Twenty-one steroids (Figure 2) with corticosteroid and testosterone binding data (Table 1).⁸⁻¹⁰ Furthermore, this set amplified by 10 additional steroids (Figure 3) with corticosteroid binding data (Table 1).⁸⁻¹⁰ (b) Thirty-seven β -carboline, pyridodiindole, and CGS compounds with affinity for the benzodiazepine receptor inverse agonist site (Table 2).^{11,12} (c) Fifteen substituted imidazoles with dissociation constant (pK_a) data (Table 3).^{12,13} (d) Forty-nine substituted benzoic acids with Hammett σ constant data (Table 4).^{12,14} (e) Thirty-three non-nucleoside reverse transcriptase HIV-1 inhibitors (NNRTI's) of the TIBO related series (Table 5), with measured inhibition of cytopathic effects of HIV-1 in MT-4 cells.¹⁵

The objective will be to show, in as direct a manner as possible, that the 13 descriptors delineated in the previous section yield predictive molecular activity in a 3D-QSAR analysis. Even though it is not necessary to perform molecular superposition, one must still address the issue of conformer choice. With respect to the 21 steroids, such a choice had been determined in the original CoMFA paper.⁸ Since all 21 structures are available from SYBYL 6.01,¹⁶ they have been used without modification in the present investigation. All other structures have been determined in a standard manner. The structures have first been generated with standard chemical bond lengths and bond angles. They were then optimized by a TRIPOS force-field¹⁷ calculation. All model building and 3D-QSAR statistical calculations have been performed with SYBYL 6.01 on an IBM Model 530H RS/6000. Rotatable bonds were

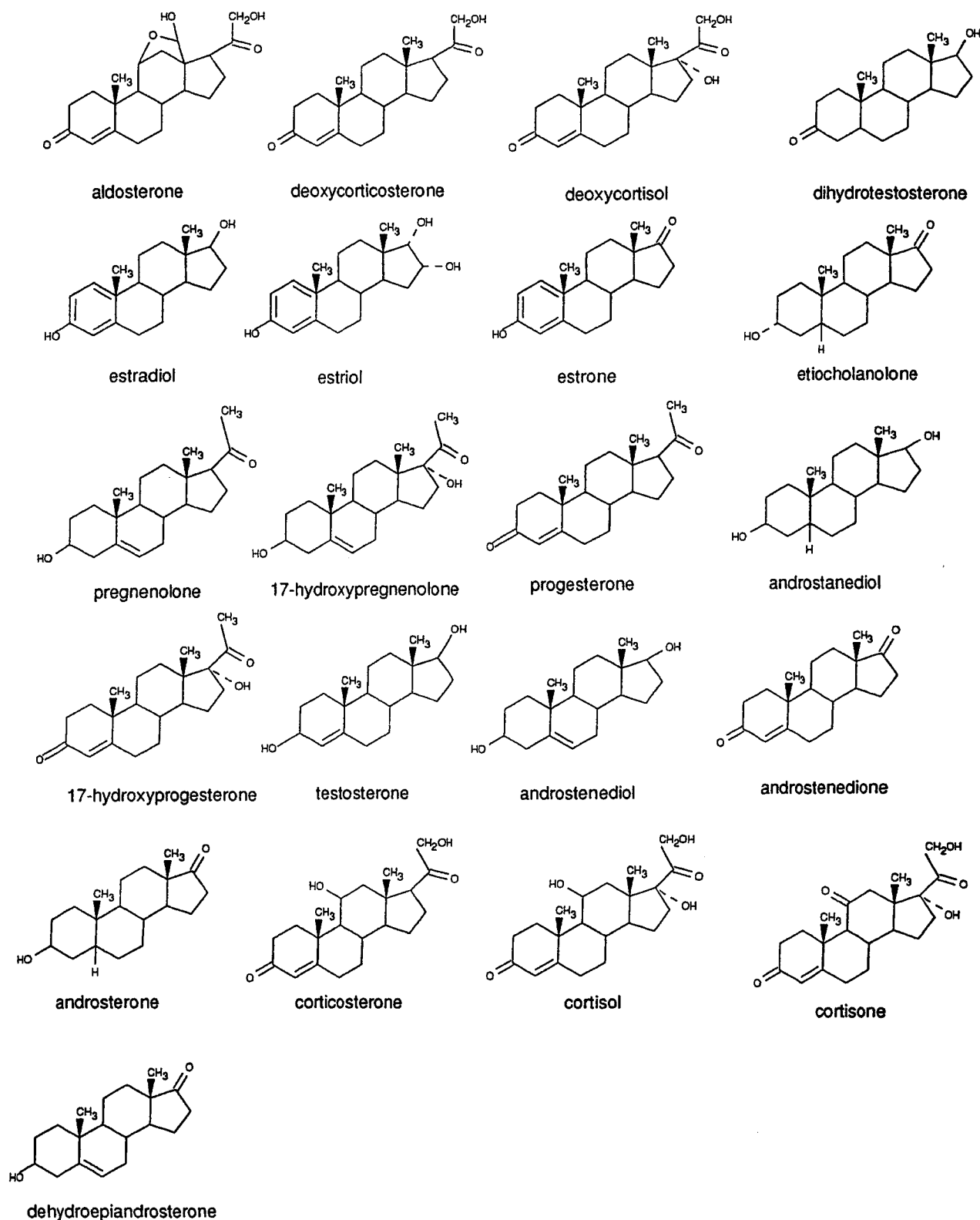


Figure 2. Twenty-one steroid structures for which corticosteroid and testosterone binding were predicted.

assigned, and a high-resolution (10° increment) systematic search was performed to identify the conformer of lowest force-field energy. This conformer was chosen and subjected to a final force-field optimization. The only set of molecules subjected to a subsequent MOPAC AM1¹⁸ geometry optimization were the 33 TIBO's,¹⁵ Table 5. Separate programs have been written and utilized for the calculation of the descriptors.

Dipole and quadrupole moments were calculated by three different procedures. Three different procedures

were utilized since, in the absence of measured dipole and quadrupole moments of the series under investigation, a measure of electrostatic moment reliability can only be achieved by observing consistency with respect to the different methods utilized. One method utilized the assignment of Gasteiger–Marsili charges¹⁹ at the atomic sites. The molecular dipole and quadrupolar components were then obtained by performing the appropriate sums over the atomic partial charges. Another similar procedure utilized Mulliken partial

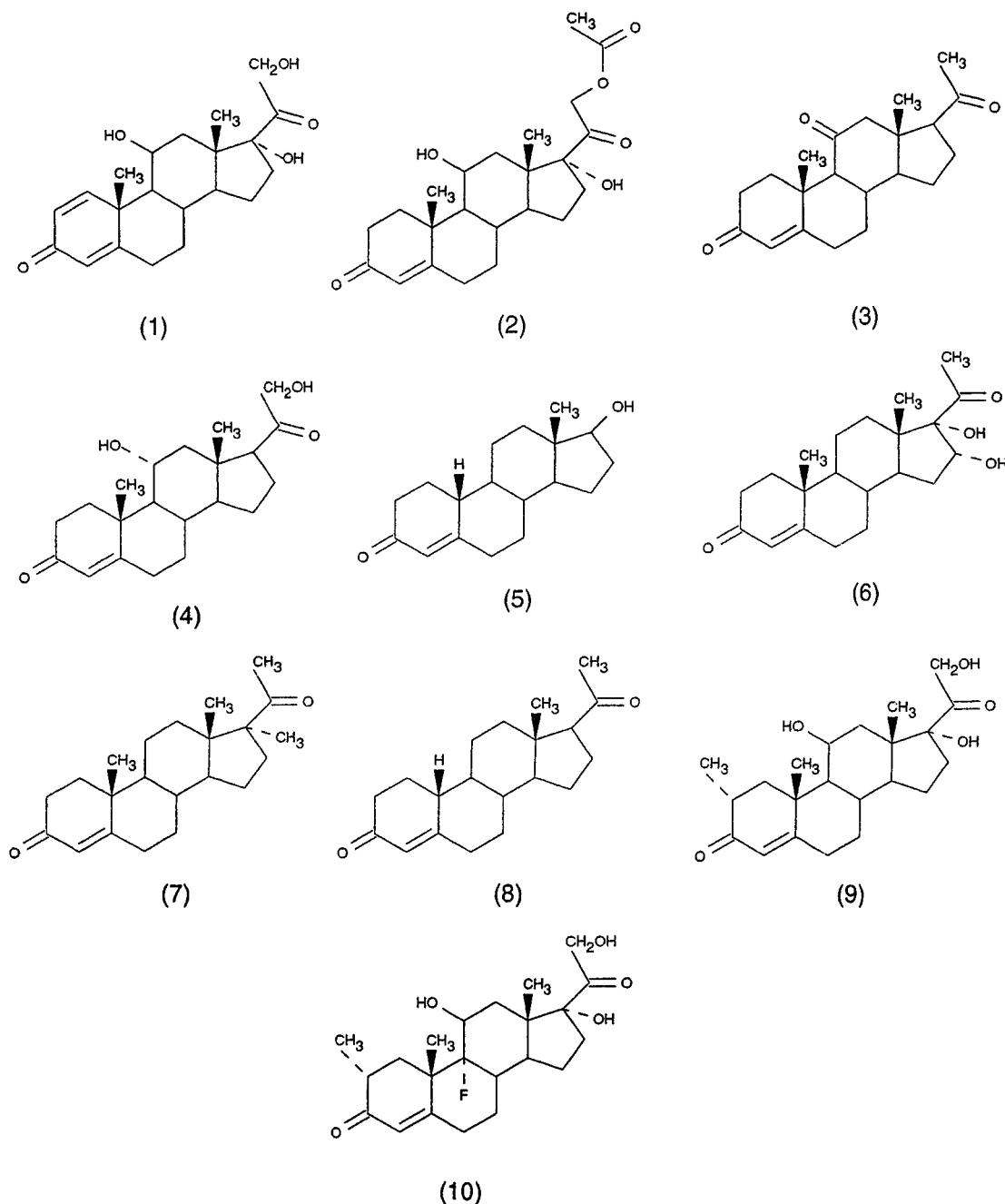


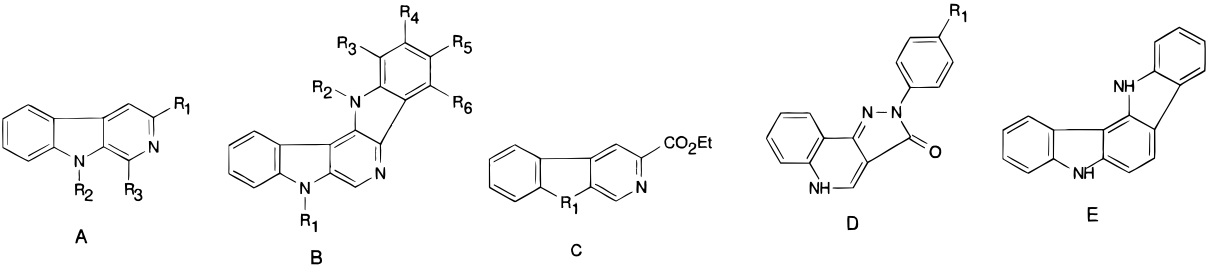
Figure 3. Ten steroid structures that form a complement with the 21 structures shown in Figure 2, yielding 31 structures for which corticosteroid binding was predicted.

charges from an AM1 MOPAC calculation.¹⁸ Finally, Gaussian 92²⁰ ab initio calculations were performed with either an STO-3G* or 6-31G* basis and the moments from the extended electronic charge distributions utilized in the calculations. For sulfur compounds, in particular, the 6-31G* split-valence polarization function basis provides significantly reduced mean absolute errors between the calculated and observed dipole moments when compared with smaller basis set approximations.²¹

The procedure for performing the 3D-QSAR studies is, therefore, straightforward. In summary, one generates the structures and chooses appropriate conformers. One then calculates the center-of-mass and determines the principal inertial components and axes for each of the conformers about this center. Utilizing the calculated dipolar and quadrupolar components for an arbitrary Cartesian frame of reference the center-of-dipole

is calculated for each conformer, and the principal quadrupolar moments and axes are obtained about this center. Dipolar, quadrupolar, and displacement descriptors are then calculated with reference to the principal inertial axes translated such that its origin is superposed on the center-of-dipole. This yields a set of 13 descriptors are enumerated in Section 4. These descriptors are then utilized in a partial least squares (PLS) analysis with the standard cross-validation "leave-one-out" procedure.

It should be emphasized that for the examples discussed in the next section, no attempt has been made to optimize the cross-validated r^2 , by manipulating the input to the PLS calculation. There has been no elimination of molecular "outliers". There has been no attempt to optimize r^2 through conformer selection and no optimization as a result of the selective choice of independent variables. The initial objective of this

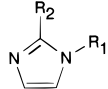
Table 2. β -Carbolenes, Pyridodiindoles, and CGS Compounds


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				-1.000
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	D	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

investigation has been simple and straightforward, namely, to show that one can perform 3D-QSAR calculations utilizing the molecular moment information described that yield predictive ability.

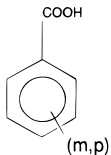
6. 3D-QSAR Examples

Figure 4 is a plot of the predicted versus the actual binding affinities for the corticosteroid binding (cbg) affinities of the 21 steroids examined in the original CoMFA study.⁸ The standard leave-one-out cross-validation PLS procedure has been utilized with the 13 internal molecular moment descriptors as independent variables. The cross-validated r^2 found is 0.674 with two optimal PLS components chosen to yield the highest r^2 . Figure 5 is a plot of the corresponding quantities for the testosterone binding (tbg) affinities of the 21 steroids. The cross-validated r^2 found is 0.412 with the optimal number of components equal to two. As previously mentioned, these structures, taken from the SYBYL 6.01 database, had Gasteiger-Marsili¹⁹ charges assigned. The electrostatic moments were calculated by performing sums over the partial charges at the

Table 3. Substituted Imidazoles with pK_a Data


no.	R ₁	R ₂	measured 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	SCH ₃	5.95

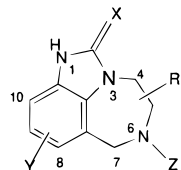
atomic sites. It is of interest that the scatter plot for the tbg binding data is less disperse than suggested by the low r^2 value. This is a consequence of the "outlier"

Table 4. Hammett Constants of Substituted Benzoic Acids


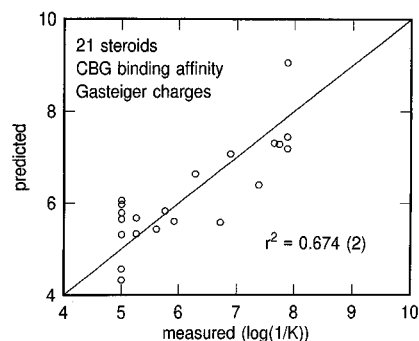
no.	substituent	Hammett constant	no.	substituent	Hammett constant
1	H	0.00	26	p-Br	0.23
2	m-Br	0.39	27	p-CF ₃	0.54
3	m-CF ₃	0.43	28	p-CH ₃	-0.17
4	m-CH ₃	-0.07	29	p-Cl	0.23
5	m-Cl	0.37	30	p-CN	0.66
6	m-CN	0.56	31	p-F	0.06
7	m-F	0.34	32	p-I	0.18
8	m-I	0.35	33	p-NH ₂	-0.66
9	m-NH ₂	-0.16	34	p-NO ₂	0.78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

with an enhanced value of predicted activity. This "outlier" is estriol, one of three of the 21 steroids with an aromatic A ring, the other two being estradiol and estrone. Examination of the in-plane orientation of the hydroxyl group coupled to this ring shows the orientation of this group to be different from the in-plane orientation of the corresponding hydroxyl groups of estradiol and estrone. Reorienting the estriol hydroxyl group so that its orientation is similar to that of estradiol and estrone yields tbg and cbg cross-correlated leave-one-out predictive r^2 values of 0.574(4) and 0.715(2), respectively. This illustrates the potential sensitivity of CoMMA with respect to variations in the directionality of local moments associated with electronegative substituents.

Whereas Gasteiger charges have been shown to qualitatively reproduce trends observed with respect to the several measured molecular dipole moments,²² one does not know how accurately such charges yield either dipolar or molecular quadrupolar moments for the class of structures we have investigated. A reasonable procedure to provide a further check on the reliability of the moments utilized by CoMMA would involve the use of molecular moments obtained independently from a number of different procedures yielding molecular charge distributions. For example, one may utilize the moments calculated directly from the spatially extended molecular orbitals obtained with a quantum chemistry program. Ab-initio calculations have been performed at the 6-31G* Hartree-Fock level of approximation for the 21 steroids and the predicted versus actual binding affinities shown in Figures 6 and 7 for cbg and tbg data, respectively. Of interest is not only that the calculations yield results indicating predictability but also that the difference between the calculated r^2 's for the cbg and

Table 5. HIV-1 Inhibitory Activity of TIBO Derivatives


no.	R	X	Y	Z	IC ₅₀ (μM)
1a	H	S	8-Cl	DMA	0.046
1b	H	S	9-Cl	DMA	0.16
1u	4-Me	O		2-MA	32.1
1v	4-Me(S)	S	9-Cl	2-MA	0.67
1w	4-Me(R)	S	9-Cl	CH ₂ -c-Pr	2.18
1z	4- <i>i</i> -Pr	O		2-MA	12.6
1ac	4- <i>n</i> -Pr	O		2-MA	48.1
1ai	7-Me	O		DMA	12.1
1aj	7-Me	O	8-Cl	DMA	0.145
1ak	7-Me	O	9-Cl	DMA	0.16
1al	7-Me	S		<i>n</i> -Pr	2.43
1am	7-Me	S		DMA	0.078
1an	7-Me	S	8-Cl	DMA	0.012
1ao	7-Me	S	9-Cl	DMA	0.023
1as	4,5-di-Me(<i>cis</i>)	O		DMA	56.7
1at	4,5-di-Me(<i>cis</i>)	S		DMA	2.22
1au	4,5-di-Me(<i>trans</i>)	S		CH ₂ -c-Pr	13.4
1az	5,7-di-Me(<i>trans</i>)	S		DMA	0.042
1ba	5,7-di-Me(<i>cis</i>)	S		DMA	1.15
1bb	5,7-di-Me(<i>R,R, trans</i>)	O	9-Cl	DMA	0.23
1bc	5,7-di-Me(<i>R,R, trans</i>)	S	9-Cl	DMA	0.48
1bg	4,7-di-Me(<i>trans</i>)	S		DMA	25.8
1bq	5-Me(S)	S	8-Cl	DMA	0.005
1bs	5-Me(S)	O	9-Cl	DMA	0.18
1bt	5-Me(S)	S	9-Cl	DMA	0.043
1bu	5-Me(S)	S	9-Cl	CH ₂ -c-Pr	0.034
1bv	5-Me(S)	S		CH ₂ -c-Pr	0.060
1bx	5-Me	O		<i>n</i> -Pr	59.7
1by	5-Me	S		<i>n</i> -Pr	1.67
1bz	5-Me	O		2-MA	34.7
1ca	5-Me	S		DMA	0.097
1cb	5-Me(S)	O		DMA	3.32
1cc	5-Me(S)	S		2-MA	0.026

**Figure 4.** Predicted vs actual binding for the corticosteroid binding (cbg) affinities of the 21 steroids shown in Figure 2. Electrostatic moments calculated from Gasteiger charges. Partial least squares (PLS) cross-validated leave-one-out r^2 's given along with the optimal number of components.

tbg binding data is qualitatively maintained, the more significant predictability associated with the cbg data. The ab-initio calculations also yield greater r^2 values for the steroids than obtained with the calculations utilizing Gasteiger charges.

Table 6 lists the results for the five different molecular series investigated. Three separate calculations have been performed for each molecule: one utilizing Gasteiger charges,^{19,22} one utilizing AM1 MOPAC Mulliken charges,¹⁸ and one utilizing the dipolar and quadrupolar moments obtained directly from the Gaussian 92²⁰ extended molecular orbital ab-initio basis. The choice

Table 6^a

method	Partial Least Squares (PLS) Cross-Validated r^2 's						
	21 steroids		31 steroids	15 imidazoles	49 benzoic acids	37 carbolines, etc.	33 TIBO's
	cbg	tbg	cbg				
Gasteiger	0.674 (2)	0.412 (2)	0.559 (1)	0.788 (2)	0.398 (1)	0.493 (4)	0.468 (1)
	-0.054	0.485	-0.041	-0.896	0.126	-0.112	-0.232
AM1 Mulliken	0.822 (2)	0.442 (2)	0.675 (8)	0.722 (2)	0.567 (5)	0.378 (2)	0.374 (2)
	0.100	0.162	0.000	-1.839	0.137	-0.115	-0.398
6-31G*/STO-3G*	0.828 (3)	0.693 (4)	0.689 (6)	0.698* (2)	0.690* (13)	0.394 (2)	0.358 (2)
	-0.101	0.028	-0.086	-0.117	0.173	-0.188	-0.395

^a An asterisk indicates that an STO-3G* basis has been utilized in the ab-initio calculation.

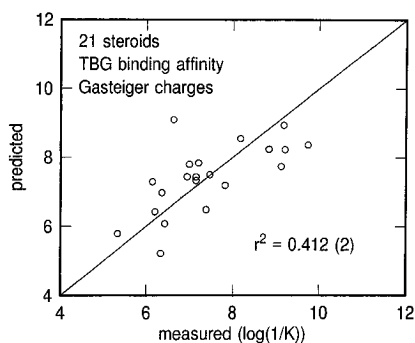


Figure 5. Predicted vs actual binding for the testosterone binding (tbg) affinities of the 21 steroids shown in Figure 2. Electrostatic moments calculated from Gasteiger charges. Partial least squares (PLS) cross-validated leave-one-out r^2 's given along with the optimal number of components.

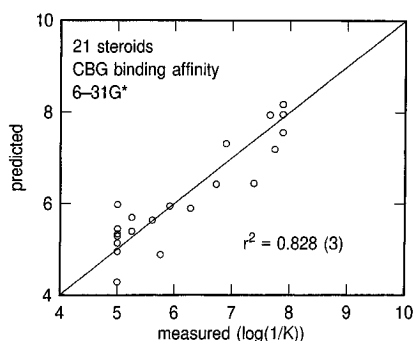


Figure 6. Predicted vs actual binding for the corticosteroid binding (cbg) affinities of the 21 steroids shown in Figure 2. Electrostatic moments obtained from 6-31G* ab-initio quantum chemistry calculations. Partial least squares (PLS) cross-validated leave-one-out r^2 's given along with the optimal number of components.

of either an STO-3G* or 6-31G* basis was dictated by the ability to use the built-in Gaussian 92 basis set for every element in the molecular series under investigation. Bromine substituents therefore precluded utilization of the 6-31G* basis for the imidazole and benzoic acid series. Results are also shown for 31 steroids, namely the previous 21 with the additional 10 (Figure 3) treated in the original CoMFA study⁸ as well in subsequent studies.^{9,10} In contrast to the other studies involving the 10 additional steroids, we have obtained r^2 for the leave-one-out cross-validated PLS procedure for the entire set of 31 molecules. All other r^2 's have also been obtained with the leave-one-out cross-validation PLS procedure. The number of optimal PLS components is listed in parentheses.

Below each entry we have also listed the r^2 obtained for a contrived "data-shift". Since CoMMA is a new method for performing 3D-QSAR and there has been concern for chance correlation using PLS,²³ we have

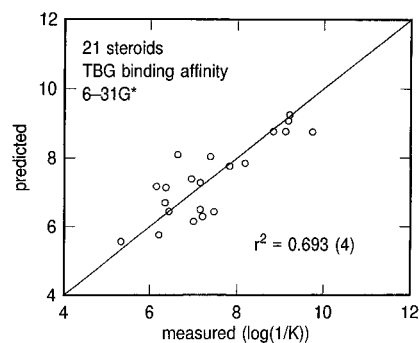


Figure 7. Predicted vs actual binding for the testosterone binding (tbg) affinities of the 21 steroids shown in Figure 2. Electrostatic moments obtained from 6-31G* ab-initio quantum chemistry calculations. Partial least squares (PLS) cross-validated leave-one-out r^2 's given along with the optimal number of components.

performed the following check with respect to all of the cross-validated r^2 's reported in Table 6. After each 3D-QSAR analysis is performed, another analysis has been performed for a set of contrived experimental data in which the first entry of the table is assigned to the last molecular structure of the table and each of the other data entries is shifted to the row above and hence registers with a different molecular structure from which the measured property corresponds. In other words, the experimental entry or dependent variable of the second row now appears as the data entry of the first row, the third entry as the data entry of the second row, and so forth until the entire molecular spreadsheet is reassigned. Such data-shift is seen to significantly diminish the correlation between predicted and measured activity in every case except one. Consequently, the values of the cross-validated PLS r^2 entries listed in Table 6 suggest that there is a significant correlation between the measured chemical/biological activities of the molecular series and the values of the CoMMA descriptors.

Fifteen imidazoles have been included in the training set instead of the 16 treated previously^{12,13} since the reported^{24,25} 2-(2-pyridyl)imidazole acid dissociation constant is for the ionic protonated species. Furthermore, for this molecular series, only 11 descriptors have been utilized, since all of the 15 molecular structures are essentially planar, the only atoms above or below the plane being hydrogen atoms associated with alkane substituents. As a consequence, the number of molecules in all of the training sets is greater in number than the number of descriptors. While standard multiple regression procedures could, therefore, have been utilized, PLS, singular value decomposition (SVD),²⁶ or some other statistical procedure is essential due the highly correlated nature of the moment descriptors. On

the other hand, the small number of descriptors provide computational turn around for the small sets of molecules examined that appears close to instantaneous.

The 49 substituted benzoic acids consist predominantly of planar species. On the other hand, there are a number electronegative substituents located on either side of the aromatic plane. Consequently, deleting descriptors in a manner analogous to that performed for the imidazoles, i.e., descriptors that signal only the presence of these nonplanar substituents was not found to change the calculated r^2 by any significant amount. The results shown in Table 6 for the substituted benzoic acids are, therefore, results obtained with the full set of 13 descriptors.

The relatively low r^2 obtained for the β -carbolines, dipyridiindoles, and CGS compounds is not surprising since modulation of these structures is mainly a consequence of moving methyl groups to different substituent locations or by the attachment of different alkyl chain lengths at a particular location. Such modifications hold little consequence for modulation of the electrostatic molecular properties. Within the β -carboline series alone, the present method yields no predictive information above that obtained by calculating the average dosage. Since this molecular series is also essentially planar, only 11 descriptors have been utilized as for the imidazole series.

The 33 TIBO molecules selected for study have been taken from Table 1 of ref 15. This table is composed of 81 molecules. The selection was determined by neglecting all molecules observed to be relatively inactive as characterized by a greater than (>) symbol in the IC₅₀ dosage column. Also neglected were all molecules with the dosage footnoted to indicate a significant spread in certain of the IC₅₀ dosage results. Furthermore, since molecular structures have not been determined for almost all of the molecules of this series, a common structural scaffold was assumed for all 33 molecules and a single conformation was adopted for the seven-membered ring. NMR studies in solution over a range of temperatures including room temperature for a TIBO precursor R78362²⁷ and for R82913²⁸ suggest conformational flexibility of this seven-membered ring. Substituent chiral designations listed in Table 1 of ref 15 were incorporated during model building. All calculations were performed for the unprotonated neutral species. With these approximations in mind for the TIBO series, it is of interest that we are able to obtain the predictive r^2 values listed in Table 6. Note that the data-shift imposed on this series destroys the predictive ability of the calculation.

One should note that the comparison of cross validated r^2 's for a particular molecular series calculated with several different charge distributions is not sufficient to guarantee consistency. It is also necessary to compare the relative importance or selectivity of the descriptors in determining the chemical/biological activity variances. We are currently investigating selectivity and consistency with respect to several different ways of obtaining electrostatic moments. Since there are no lists of measured moments for the types of molecular series we are investigating, a strategy to suggest validity of the moment calculations will be based upon consistency, i.e., the ability of the different moment calculations to yield similar sets of descriptors predomi-

Table 7. Quadrupolar Descriptors^a

molecule	pK _a	q _{xx}	q _{yy}
1	3.82	0.837	0.205
2	2.30	0.640	0.042
3	7.12	0.135	1.247
4	8.54	0.663	0.042
5	-0.48	-0.244	-0.576
6	3.79	0.898	0.153
7	3.55	0.772	0.161
8	7.73	1.228	0.038
9	2.40	0.681	0.044
10	6.99	1.338	0.114
11	7.86	1.344	0.017
12	8.46	1.922	0.377
13	-0.81	-0.408	-0.669
14	6.48	2.794	0.000
15	5.95	2.229	0.001

^a Quadrupolar units are Å².

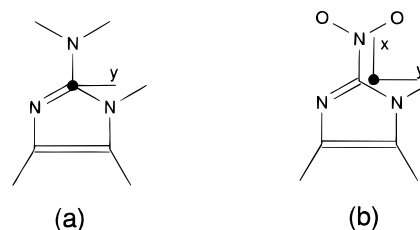


Figure 8. Amine- (a) and nitro-substituted (b) imidazoles. Filled in circle represents the location of the center-of-dipole.

nantly responsible for the observed biological or chemical activity variances.

In connection with issues of selectivity and consistency of descriptors, we will describe initial results obtained for the 15 imidazoles. Utilizing only the five descriptors q_{xx} , q_{yy} , I_x , I_y , and I_z obtained from the STO-3G* calculations yields a predictive cross-validated r^2 value of 0.772(2), an increase over the value of 0.698(2) in Table 6 obtained with all of the 11 descriptors of the imidazole series. This is the maximum that can be achieved by variable selectivity among the 11 descriptors. The AM1 MOPAC calculations similarly select these variables as important. Utilizing only the two descriptors q_{xx} and q_{yy} yields a predictive cross-validated r^2 of 0.690(2), which indicates that these two descriptors are responsible for the major correlation with the observed pK_a of this molecular series. Table 7 lists the observed pK_a and the two descriptor values for this molecular series. One notes a correspondence between the pK_a values and values of the descriptors, namely, the larger the value of pK_a, the larger the values of the descriptors. The amine- and nitro-substituted imidazoles, molecules **12** and **13**, illustrate an interesting case in contrast. Figure 8 shows a diagram of these molecular structures along with the inertial axes translated so that the origin is at the center-of-dipole. The center-of-dipole is shown as a small filled-in black circle in the figure. The center-of-dipole of each molecule is either at or near the ring carbon to which the substituent is bound. The z direction of the inertial axes is into the paper. The x inertial axis with the amine substituent is along the bond between the ring carbon atom and substituent nitrogen atom. The negative value of the quadrupolar moments for the nitro substituent is a result of the extensive delocalization of electronic charge about this center and is directly a consequence of the electron-withdrawing character of the nitro substituent and the presence of the oxygen atoms at distal locations

Table 8

	CoMMA	pre CoMMA
21 steroids, cbg	0.828 (3)	0.686 (3)
21 steroids, tbg	0.693 (4)	0.399 (2)
31 steroids	0.689 (6)	0.333 (1)
imidazoles	0.698 (2)	0.258 (3)
benzoic acids	0.690 (13)	0.472 (1)
carbolines, etc.	0.394 (2)	0.448 (3)
TIBO's	0.358 (2)	0.314 (1)

from the center-of-dipole. The large positive values of the moments arising with the amine substituent are directly related to the proximity of donated electron charge to the center-of-dipole.

Finally, since we have not only pointed out that quadrupolar descriptors complete the set of mass and charge descriptors up to and including all moments of second order, but have tacitly suggested that sets of descriptors based upon such moments require the second order quadrupolar moments to yield significant 3D-QSAR predictability, we will show results obtained with and without these descriptors. Table 8 lists predictive r^2 values for the seven different sets of results already listed in Table 6. We have reproduced the predicted cross-validated results obtained with moments obtained only from the ab-initio calculations. Presumably, these are the most reliable. Another column is shown and labelled Pre CoMMA, namely, results that one might have obtained with information not requiring the new identification² of center-of-dipole and related quadrupolar values and axes. Prior to this identification, six descriptors, namely, q , q_{xx} , q_{yy} , d_x , d_y , and d_z , could not have been defined. So the Pre CoMMA calculations utilize only the seven descriptors p_x , p_y , p_z , p , I_x , I_y , and I_z . One notes the significant deterioration in nearly every case. Again, it is of interest that the carboline, dipyrindiodiindoles, and CGS compounds appear relatively unaffected for reasons previously alluded to. On the other hand, selecting important variables from the 11, inclusive of the quadrupolar descriptors, yields a predictive cross-validated r^2 of 0.492(7) for this series.

7. Discussion

One might ask: In what category, if any, might CoMMA be placed with respect to other 3D-QSAR procedures? Since procedures such as CoMFA, COMPASS, and molecular similarity matrices involve a detailed local description of molecular steric as well as electrostatic features, perhaps CoMMA might be assigned a category intermediate to a global scalar Hansch type QSAR or other such procedure and the 3D-QSAR procedures detailing the local features. Perhaps it might be considered a global 3D-QSAR procedure.

Since previous 3D-QSAR procedures have focused on detailed local molecular features and presumably yielded specific spatial information concerning molecular regions intimately involved in ligand-receptor binding, one might ask why CoMMA works at all since all of this very important and relevant information is not dealt with explicitly. While we cannot, at present, answer this question in the depth to which it deserves to be answered, one might suggest the following. First, utilizing inertial, dipolar, and quadrupolar moments incorporates, at some very fundamental level, information concerning how a molecule will respond to applied

torques as well as to uniform and gradient electric fields at the receptor site during the process of docking. For example, it has been suggested that the flexible polar side chains at the entrance to the non-nucleoside inhibitor binding pocket of nevirapine, TIBO R86183, and other nnrti's²⁹ may help steer the inhibitor molecule into the pocket. One might expect such incorporation of the ligand into the receptor site to be dependent upon the spatial relationship between the global molecular shape and charge distribution of the ligand. CoMMA descriptors incorporate the relationship between such features concisely. Second, even though explicit detailed local spatial information relating to ligand-receptor binding is not explicitly present in the CoMMA descriptors, such information is implicitly present in an averaged manner.

Finally, since CoMMA provides two Cartesian reference frames, namely, the principal inertial and quadrupolar axes, one might utilize either of these frames to reference local molecular features and therefore eliminate the necessity of molecular superposition in developing descriptors for a local 3D-QSAR analysis. In particular, one might reference the CoMFA grids of different molecules to their respective principal inertial frames. This would be, in some sense, the analog of structural superposition, whereas referencing local descriptors to the principal quadrupolar axes would be, in some sense, the analog of a field-fit superposition.

8. Conclusions

The primary objective of the present paper has been to demonstrate that internal molecular 3D shape and charge descriptors can be defined that enable rapid and predictive 3D-QSAR analyses without the requirement of molecular superposition, i.e., without requiring a comparison between different molecules. Such descriptors utilize spatial moments of the molecular mass and charge distributions up to and including second order as well as related quantities. The ability to calculate the second-order moments of the charge distribution of neutral molecules for the purpose of comparison depends critically upon the ability to define an appropriate center of electrostatic multipolar expansion. Finally, it will be of interest to determine the utility of such small concise set of descriptors in addressing issues of large-scale screening as well as molecular diversity. Future investigations should lead to a clearer understanding of the utility, strengths, and weaknesses of the proposed method.

Acknowledgment. We would like to thank Tom Jackman and Mike Pitman for several valuable discussions and Tom Jackman for assistance with several of the calculations.

References

- (1) Sheridan, R. P.; Miller, M. D.; Underwood, D. J.; Kearsley, S. K. Chemical Similarity Using Geometric Atom Pair Descriptors. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 129–136.
- (2) Platt, D. E.; Silverman, B. D. Registration, Orientation and Similarity of Molecular Electrostatic Potentials through Multipole Matching. *J. Computat. Chem.* **1996**, *17*, 358–366.
- (3) See, for example: Goldstein, H. *Classical Mechanics*, 2nd ed.; Addison Wesley Publishing Co.: New York, 1980.
- (4) See, for example: Jackson, J. D. *Classical Electrodynamics*, 2nd ed.; John Wiley and Sons: New York, 1975.
- (5) Buckingham, A. D. Permanent and Induced Molecular Moments and Long Range Intermolecular Forces. In *Advances in Chemical Physics Volume 12*; Hirschfelder, J. O., Ed.; 1967; p 107.

- (6) Inami, Y.; Tomita, T.; Terada, Y. Quantitative Structure-Activity Relationship Analysis of Phencyclidine Derivatives. I. *Chem. Pharm. Bull.* **1991**, *39*, 1426–1429.
- (7) Cardozo, M. G.; Iimura, Y.; Sugimoto, H.; Yamanishi, Y.; Hopfinger, A. J. QSAR Analyses of the Substituted Indanone and Benzylpiperidine Rings of a Series of Indanone-Benzylpiperidine Inhibitors of Acetylcholinesterase. *J. Med. Chem.* **1992**, *35*, 584–589.
- (8) 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.
- (9) Good, A. C.; Sung-Sau, S.; Richards, W. G. Structure-Activity Relationships from Molecular Similarity Matrices. *J. Med. Chem.* **1993**, *36*, 433–438.
- (10) Jain, A. N.; Koile, K.; Chapman, D. Compass: Predicting Biological Activities from Molecular Surface Properties. Performance Comparisons on a Steroid Benchmark. *J. Med. Chem.* **1994**, *37*, 2315–2327.
- (11) Allen, M. S.; Tan, Y.; Trudell, M. M.; Narayanan, K.; Schindler, L. R.; Martin, M. J.; Schultz, C.; Hagen, T. J.; Koehler, K. F.; Coddling, P. W.; Skolnick, P.; Cook, J. M. Synthetic and Computer-Assisted analyses of the Pharmacophore for the Benzodiazepine Receptor Inverse Agonist Site. *J. Med. Chem.* **1990**, *33*, 2343–2357.
- (12) Good, A. C.; Peterson, S. J.; Richards, W. G. QSAR's from Similarity Matrices. Technique Validation and Application in the Comparison of Different Similarity Evaluation Methods. *J. Med. Chem.* **1993**, *36*, 2929–2937.
- (13) Kim, K. H.; Martin, Y. Direct Prediction of Dissociation Constants (pKa's) of Clonidine-like Imidazoles, 2-substituted Imidazoles, and 1-Methyl-2-substituted-imidazoles from 3D structures Using a Comparative Molecular Field Analysis (CoMFA) Approach. *J. Med. Chem.* **1991**, *34*, 2056–2060.
- (14) Kim, K. H.; Martin, C. M. 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.
- (15) Breslin, H. J.; Kukla, M. J.; Ludovici, D. W.; Mohrbacher, R.; Ho, W.; Miranda, M.; Rodgers, J. D.; Hitchens, T. K.; Leo, G.; Gauthier, D. A.; Ho, C. Y.; Scott, M. K.; De Clercq, E.; Pauwels, R.; Andries, K.; Janssen, M. A. C.; Janssen, P. A. J. Synthesis and Anti-HIV-1 Activity of 4,5,6,7-Tetrahydro-5-methylimidazo-[4,5,1-jk][1,4]benzodiazepin-2(1H)-one (TIBO) Derivatives. 3. *J. Med. Chem.* **1995**, *38*, 771–793.
- (16) Available from TRIPOS Associates Inc., 1699 S. Hanley Rd., St. Louis, MO.
- (17) Clark, M.; Cramer, R. D. III; Van Opdenbosch, N. Validation of the Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
- (18) Stewart, J. J. P. MOPAC: A Semiempirical Program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1–105.
- (19) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3288.
- (20) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *Gaussian 92, Revision C*; Gaussian Inc.: 4415 Fifth Ave., Pittsburgh, PA 15213, 1992.
- (21) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *AB Initio Molecular Orbital Theory*; John Wiley and Sons: New York, 1986.
- (22) Gasteiger, J.; Guillen, M. D. Dipole Moments Obtained by Partial Equalization of Orbital Electronegativity. *J. Chem. Res. Synop.* **1983**, 304–305.
- (23) Clark, M.; Cramer, R. D. III. The Probability of Chance Correlation Using Partial Least Squares (PLS). *Quant. Struct.-Act. Relat.* **1993**, *12*, 137–145.
- (24) Catalan, J. Basicity and Acidity of Azoles. *Adv. Heterocycl. Chem.* **1987**, *41*, 187–274.
- (25) Eibeck, W. J.; Holmes, F. Heterocyclic Chelating Agents. Part II. Metal Complexes of 2-(2-Pyridyl)imidazole. *J. Chem. Soc. A* **1967**, 1777–1782.
- (26) Press, W. H.; Flannery, B. P.; Teukolsky, W. T. *Numerical Recipes: The Art of Scientific Computing*; Cambridge University Press: Cambridge, 1986; p 52.
- (27) Caldwell, G. W.; Gauthier, A. D.; Kukla, M. J. Conformational Analysis of (S)-4,5,6,7-tetrahydro-5-methylimidazo[4,5,1-jk][1,4]-benzodiazepin-2(1H)-one (R78362). *Spectrosc. Lett.* **1993**, *26* (6), 1005–1022.
- (28) Caldwell, G. W.; Gauthier, A. D.; Leo, G. C.; Kukla, M. J. Conformational Analysis of the Reverse Transcriptase Inhibitor (+)-(S)-4,5,6,7-tetrahydroimidazo-9-chloro-5-methyl-6-(3-methyl-2-butenyl)imidazo[4,5,1-jk][1,4]-benzodiazepin-2(1H)-thione (TIBO; R82913). *Tetrahedron Lett.* **1993**, *34* (13), 2063–2066.
- (29) Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A. J.; Hughes, S. H.; Arnold, E. Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Struct. Biol.* **1995**, *2* (5), 407–415.

JM950589Q