Three-Dimensional Moments of Molecular Property Fields

B. David Silverman[†]

IBM Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598

Received June 13, 2000

Descriptors that capture certain three-dimensional molecular features and do not require molecular superposition or alignment for the assignment of molecular similarity have recently been proposed and investigated. Among these, moments of molecular features have been utilized in the QSAR of several molecular series. The present work examines certain formal aspects of the use of moment expansions for which the zero-order moment of a property field is nonvanishing. The first-order term of such moment expansion is then dependent upon the origin of expansion, and it is pointed out that expansion about the molecular centroid is descriptive of first-order differences about the property-field mean. For a hydrophobic property field, this first-order term is just the Eisenberg hydrophobic moment. Second-order moments about the centroid can be written as components of the WHIM covariant matrix. Moment expansions are also performed about the property-field center in analogy with expansions about the center of mass. For such expansion, a set of descriptors consisting of moments of the molecular density and of a molecular hydrophobic property field as well as of related quantities are used in a QSAR of the binding of 74 polyhalogenated aromatic molecules to the Ah cystolic receptor. This QSAR has been named CoMMA2 to distinguish it from CoMMA, for which the zero-order moment of the expansion vanishes.

1. INTRODUCTION

The three-dimensional characterization of molecular physical and chemical properties has been a subject of interest in connection with the numerous procedures that attempt to correlate such characterization with molecular biological activity. Since one's a priori expectation is that three-dimensional molecular features should be central to the delivery and binding of a drug molecule to its targeted receptor site, one expects such characterization to augment the classic Hansch[1] or topological[2] type molecular descriptors.

Many of the three-dimensional procedures such as CoMFA[3] as well as CoMFA-related procedures[4−7] involve the detailed enumeration of molecular properties over a set of grid points and subsequently require an alignment or superposition step prior to the assignment of molecular similarity. That is the price that one must pay by requiring the detailed characterization of a three-dimensional molecular property field, whether it is steric, electrostatic, or hydrophobic. There have, however, been a number of characterizations[8−12] dependent upon three-dimensional structure that capture molecular features in ways not requiring an alignment or superposition step for the assignment of molecular similarity. The alignment-free procedures also generate a relatively small set of three-dimensional descriptors, thus enabling greater ease of statistical analysis. While similarity matrices[13,14] significantly reduce the number of descriptors compared with the grid-based procedures, they still require a molecular alignment step.

Molecular moments descriptive of some molecular property provide a small set of alignment-free descriptors that can be utilized in a quantitative structure−activity relationship (QSAR). We had previously examined moments of the shape and charge distributions of neutrally charged molecules.[15] The molecular charge distribution is responsible for the electrostatic field external to the molecule, and a particular moment representation of the charge distribution was developed that utilized a special feature of this electrostatic field. This led to the definition of the "center of dipole", about which quadrupole descriptors had been obtained. This procedure required that the zero-order moment or net molecular charge was identically equal to zero, by definition, a condition satisfied by neutrally charged molecules.[16]

If the zero-order moment of the property field does not vanish, the nature of the expansion changes. For this case, neither the first- nor second-order moments are invariant with respect to the choice of the origin of the expansion. For such expansion, the first-order or linear moment is generally nonvanishing. Linear moments of the hydrophobic property fields of α-helical secondary structures have provided a measure of the amphiphilicity of such helices.[17] This has been used in identifying the helical regions of proteins that bind to the surface of biological membranes. One might expect such moments to play a useful role in the characterization of molecules. The next order in the expansion about the centroid of the molecule yields second-order moments which can be written as the elements of the WHIM covariance matrix.[8,9]

Utilizing the centroid as a reference origin implicitly assumes the existence of a distribution of molecular density. The centroid is defined as the origin of the moment expansion for which the first-order moment of the molecular density distribution vanishes. In other words, it is the property-field center of this density distribution. This distribution is generally simply chosen by assigning unit property values

[†] E-mail: silverma@us.ibm.com. FAX: (914) 945-4104. Phone: (914) 945-1135.

10.1021/ci000457s CCC: $19.00
Published on Web 09/20/2000

at each atomic center. Diagonalization of a three-dimensional second-rank tensor composed of the second-order moments of this distribution will yield moments of geometry as principal components associated with each of the principal geometric axes, in analogy with the moments of inertia and inertial axes, respectively. The moments of geometry, so obtained, will be used as molecular shape descriptors in the present paper. It should be realized, however, that the molecular density distribution could have been chosen differently, e.g., by a series of points or by a continuous function on either of the Connolly or van der Waals surfaces.

Since moment expansions can be performed about an arbitrary origin, one can expand property fields about the property-field center in analogy with expansions performed about the center of mass. The property-field center is obtained by translating the origin of expansion to the location about which the first-order property-field moment vanishes. Information that had been provided by the first-order moment in the expansion about the centroid is now provided by the displacement between the two centers, namely between the centroid and the property-field center. Diagonalization of the second-order property-field second-rank-tensor yields a set of eigenvalues that can be used as descriptors to characterize the particular property field.

Section 2, which traces out these ideas in some detail, makes reference to two different distributions, one of the property field of molecular density and the other of an arbitrary molecular property field. Both yield separate moment expansions. Each expansion about its respective property-field center yields a set of principal axes. A description of the explicit relationship between two such distributions can be obtained in a number of different ways. In the present paper such relationship is obtained by writing certain second-order moments of the property field in the frame of the principal axes provided by the molecular density distribution. Principal axes of each of the distributions provide the capability of relating the distributions as well as the capability of comparing moments of third or higher order within each of the distributions.

Section 2, Molecular Moments, describes the formal development of the property-field moments and the selection of the CoMMA2 descriptors. Section 3, Comparative Molecular Moment Analysis (CoMMA2): The Hydrophobic Property field of 74 Polyhalogenated Aromatic Compounds, examines the correlation obtained between the hydrophobic property-field moments and related quantities of these molecules to their binding at the cystolic Ah receptor site. Section 4, Discussion, provides a summary plus conclusions.

The primary objective of the present paper is to provide a detailed examination and description of certain formal aspects of the property-field moment expansions.

2. MOLECULAR MOMENTS

If $a(\vec{r})$ is a defined scalar molecular property-field density that can adopt positive or negative values over the three-dimensional molecular density distribution, $s(\vec{r})$, the zero- and first-order moments can be written as

$$m_0 = \int a(\vec{r})\, s(\vec{r})\, \mathrm{d}\vec{r} \tag{2.1}$$

$$\vec{m}_1 = \int a(\vec{r})\, s(\vec{r})\vec{r}\, \mathrm{d}\vec{r} \tag{2.2}$$

Moments of the molecular density distribution are obtained for a property-field density $a(\vec{r})$ set equal to unity. Consequently the location of the molecular centroid is given by

$$\vec{r}_c = \frac{\int \vec{r}\, s(\vec{r})\, \mathrm{d}\vec{r}}{\int s(\vec{r})\, \mathrm{d}\vec{r}} \tag{2.3}$$

The property-field distribution is arbitrary; however, it is assumed that these, as well as subsequent integrals or sums designated, are well-defined.

The functional form of $s(\vec{r})$ could be chosen to delineate a Connolly or van der Waals surface. For such choice, the property-field values, $a(\vec{r})$, would be mapped to such surface. The example currently developed, however, assigns property-field values, a_i, to the atom centers only. $s(\vec{r})$ is then the set of Dirac delta functions, δ, centered at each of the atomic sites, i, and the total number of atoms, n, composing the molecule can be written

$$n = \int s(\vec{r})\, \mathrm{d}\vec{r} = \int \sum_i \delta(\vec{r} - \vec{r}_i)\, \mathrm{d}\vec{r} \tag{2.4}$$

The zero- and first-order property-field moments are then written

$$m_0 = \sum_i a_i \tag{2.5}$$

$$\vec{m}_1 = \sum_i a_i \vec{r}_i \tag{2.6}$$

m_0 will be called the property-field weight. Therefore, if the property-field values are the atomic masses, m_0 will be the molecular weight. If the property-field values are, for example, the Ghose−Crippen[18] atomic hydrophobicity values, m_0 will be the logarithm of the partition function, Log P.

Since the zero-order moment of the expansion is nonvanishing, $\vec{m}_1$, the first-order moment, is not invariant to the choice of origin. Eisenberg[17] proposed the use of such a moment in connection with amino acid hydrophobic property values for the determination of the amphipathicity of the α-helices of membrane-bound proteins. Vector components were calculated from the α-carbon to the centroid of the residue under consideration or from the axis of the helix to the α-carbon of the residue. This provided a net vector essentially perpendicular to the helical axis. It was also stated[17] that one might use the following hydrophobic vector, invariant with respect to the choice of origin of expansion, to provide a component along the helical axis as well.

$$\vec{m}_1 = \sum_i (a_i - \bar{a})\vec{r}_i \tag{2.7}$$

with $\bar{a}$ equal to what will be called the property-field mean.

$$\bar{a} = \frac{1}{n}\sum_i a_i = \frac{m_0}{n} \tag{2.8}$$

Such a vector is then descriptive of first-order deviations about $\bar{a}$, the average over the n amino acid residue helix.

This is equivalent to referencing the property-field vector to the centroid of the structure, $\vec{r}_c$, namely

$$\vec{m}_1 = \sum_i a_i(\vec{r}_i - \vec{r}_c) \tag{2.9}$$

with

$$\vec{r}_c = \frac{1}{n}\sum_i \vec{r}_i \tag{2.10}$$

The vector, $\vec{m}_1$, will have a component perpendicular to the helical axis as well as one along the axis. As a molecular descriptor, such a vector would therefore characterize the first-order difference in property-field distribution about the property-field mean.

The integral representation of the moments in (2.1) and (2.2) that referenced the property-field distribution, $a(\vec{r})$, to the molecular density distribution, $s(\vec{r})$, enables one to retain the equivalence between the integral representations of (2.7) and (2.9). Without such mapping this equivalence would no longer hold.

Second-order moments of the property field about the centroid can be defined as components of the following second-rank tensor:

$$\tilde{m}_2^c = \sum_i a_i(\tilde{I}|\vec{r}_i - \vec{r}_c|^2 - (\vec{r}_i - \vec{r}_c)(\vec{r}_i - \vec{r}_c)) \tag{2.11}$$

$\tilde{I}$ is the unit dyadic. The superscript "c" indicates that the moment is calculated about the molecular centroid.

Had we chosen the second-rank tensor of the form

$$\tilde{m}_2^c = \sum_i a_i(\vec{r}_i - \vec{r}_c)(\vec{r}_i - \vec{r}_c) \tag{2.12}$$

it would be the same as the WHIM covariance matrix,[8] aside from a normalizing factor. The principal axes of (2.11) and (2.12) are identical; however, the eigenvalues of (2.12) are linear combinations of the moments of geometry, the eigenvalues of (2.11), that are the moments of inertia for a molecule with property-field values of unit mass assigned at each of the atomic sites. Since the eigenvalues of (2.11) and (2.12) are linearly related, they will yield identical results when used as descriptors in a multilinear regression.

Property-field descriptors can also be obtained by moment expansions about the property-field center. Early WHIM descriptors were obtained by an expansion about such an origin.[8] Moments of inertia would be obtained as the eigenvalues of (2.11) for the property field that assigned the atomic mass to the atom centers and the expansion performed about the center of mass, namely, about the property-field center of the mass distribution. More generally, the choice of (2.11) assigns the eigenvalues of all property fields as radial property-field distributions that are normal to each of the corresponding principal axes. This is, again, in analogy with the relationship between the moments of inertia and their corresponding principal axes.

Displacement to the property-field center, $\vec{a}$, from an arbitrary location is given by

$$\vec{a} = \frac{\sum_i a_i \vec{r}_i}{\sum_i a_i} \tag{2.13}$$

The property-field values need not be positive; however the property-field weight, namely the zero-order moment of the expansion, must differ from zero. For property values of varying sign, this center might not be proximate to the molecule. Furthermore, as the property-field weight goes to zero, the distance of the property-field center from any arbitrary origin at which the calculation is performed will increase without limit. This is similar to the behavior of the location of the center of dipole as the dipole of a neutrally charged molecule becomes vanishingly small.[15] The property-field center of molecular charge, namely, the "center of charge", has been previously utilized[19−23] and has provided a useful reference in distinguishing the electrostatic from the inertial properties of an ion.

The first-order property-field moments then vanish about the property-field center, and one can define second-order moments of the property field about this center that comprise the following second-rank tensor.

$$\tilde{m}_2^p = \sum_i a_i(\tilde{I}|\vec{r}_i - \vec{a}|^2 - (\vec{r}_i - \vec{a})(\vec{r}_i - \vec{a})) \tag{2.14}$$

The superscript "p" indicates that the moments are calculated about the property-field center.

Just as one can utilize the moments of geometry, I_x^g, I_y^g, I_z^g, as descriptors of molecular shape, the property-field eigenvalues of (2.14), can be utilized as descriptors of the molecular property-field distribution. For property-field values at the atomic sites that are either positive or negative, the matrix (2.14) will be real and symmetric, and hence Hermitean which guarantees real eigenvalues, Θ_1, Θ_2, Θ_3.

The magnitude of the displacement between the molecular centroid and property-field center provides one further molecular descriptor. This descriptor incorporates information that had apparently been lost in the expansion about the property-field center. This displacement, $\vec{d}$, can be written

$$\vec{d} = \vec{a} - \vec{r}_c = \vec{m}_1/m_0 \tag{2.15}$$

Therefore, information provided by the first-order moment, $\vec{m}_1$, in the expansion about the molecular centroid is retained by including the displacement, $\vec{d}$, as a descriptor for expansions performed about the property-field center for which the first-order moment vanishes. For the property field of molecular charge, the displacement of the center of charge with respect to the center of mass has assisted in the prediction of the electrophoretic mobilities of peptides.[24]

Summarizing the previous discussion: moments of the property field about the property-field center then yield the following five molecular descriptors:

$$m_0 \quad |\vec{d}| \quad \Theta_1 \quad \Theta_2 \quad \Theta_3$$

with $|\vec{d}|$ the magnitude of $\vec{d}$.

Moments have been used within the context of the spatial pattern recognition of two- and three-dimensional objects.[25,26] In principle, one is not limited to a consideration of only the first three lowest order moments for the purposes of comparison. The principal axes of (2.14) provide a reference set of axes for calculation of and comparison between higher order moments. It is of interest to see how image resolution increases with the inclusion of the moments of order higher than second.[25]

Shape descriptors of the molecule can also be included by adding the molecular geometric moments, I_x^g, I_y^g, I_z^g. Diagonalization of the second-rank property-field tensor composed of moments of the molecular density distribution, $s(\vec{r})$, provides the set of principal geometric axes that can be used to reference other property-field vectors or tensor components to molecular shape. One set of such moment descriptors that are invariant to the sensing of the principal geometric axes are the diagonal components of the second-order property-field tensor written in the principal geometric frame, namely, Q_{xx}^c, Q_{yy}^c, Q_{zz}^c.

$$Q_{xx}^c = \sum_i a_i((y - y_c)^2 + (z - z_c)^2) \qquad (2.16)$$

$$Q_{yy}^c = \sum_i a_i((x - x_c)^2 + (z - z_c)^2) \qquad (2.17)$$

$$Q_{zz}^c = \sum_i a_i((x - x_c)^2 + (y - y_c)^2) \qquad (2.18)$$

The origin of the coordinate system of (2.16)−(2.18) is chosen at the molecular centroid as indicated by the superscript "c", and the axes are aligned with the principal geometric axes. Other vector and tensor components can be written that would not be invariant to the sensing of the axes unless magnitudes of the components were used or some additional referencing was invoked to provide the sensing.

So recapitulating, the set of 11 descriptors that will be used in the calculations described in the next section are

$$I_x^g \quad I_y^g \quad I_z^g \quad m_0 \quad |\vec{d}| \quad Q_{xx}^c \quad Q_{yy}^c \quad Q_{zz}^c \quad \Theta_1 \quad \Theta_2 \quad \Theta_3$$

This set of descriptors will be named CoMMA2 to distinguish it from the CoMMA set developed for the charge distribution of neutrally charged molecules.[15]

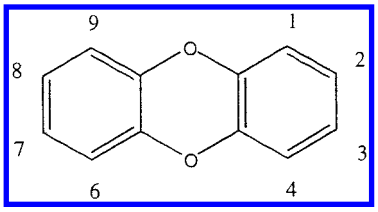
3. COMPARATIVE MOLECULAR MOMENT ANALYSIS (COMMA2): THE HYDROPHOBIC PROPERTY FIELD OF 74 POLYHALOGENATED AROMATIC COMPOUNDS

The hydrophobic atom fragment assignments of Ghose and Crippen[18] will be used in a QSAR analysis of the binding affinity of 74 polyhalogenated aromatic compounds to the Ah cystolic receptor. The moments provide a simple three-dimensional description of the property field of molecular hydrophobicity and expand this important scalar field into the realm of three dimensions. This particular set of molecules has been chosen since a number of previous QSAR studies have been performed[27−29] with the original binding data.[30] Furthermore, the calculation of Log *P* from atom fragment assignments has had a long and interesting history.[31,32] Subsequent to the introduction of CoMFA, three-dimensional hydrophobic fields were introduced[4] to supplement the steric and electrostatic interactions of the original CoMFA implementation. The three-dimensional hydrophobic moments and related quantities of the present paper might be thought of as an intermediate characterization of hydrophobicity; intermediate with respect to the characterization by one single number such as Log *P* and to the more detailed spatial three-dimensional field mapping of HINT.[4]

The 74 molecules (Tables 1−3) consist of three different series, namely, 25 polychlorinated and polybrominated dibenzo-*p*-dioxins, 35 polychlorinated dibenzofurans, and 14 polychlorinated biphenyls. The structures were determined by the TRIPOS Sybyl force field.[33] Ghose−Crippen fragment assignments replaced the SYBYL atom type assignments in the mol2 files, and a MATLAB program[34] was used to calculate the molecular moment descriptors of each of the molecules. The biphenyls were the only series of molecules that were nonplanar. Table 4 lists the calculated descriptor values for the set of 25 dibenzo-*p*-dioxins.

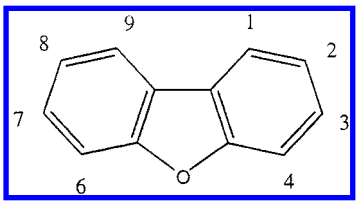
Regression[35] with no validation yields $r^2 = 0.820$ with $F = 25.8$. Interestingly, the only regression coefficient with a 95% confidence interval that does not cross zero is the coefficient of the fourth descriptor, m_0, or the calculated Log *P* value. On the other hand, the Log *P* descriptor alone yields $r^2 = 0.135$ with $F = 11.3$.

Table 1. Binding Affinities of the Polyhalogenated Dibenzo-*p*-dioxins



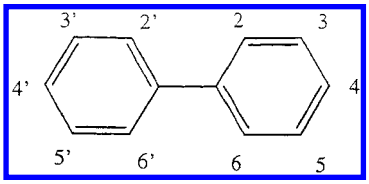
molecule	pEC_{50}	molecule	pEC_{50}
1-Cl	4.000	2-Br	6.530
1,2,3,4-Cl_4	5.886	2,3,6-Cl_3	6.658
1,2,3,4,6,7,8,9-Cl_8	5.000	2,3,6,7-Cl_4	6.796
1,2,3,4,7-Cl_5	5.194	2,3,7-Br_3	8.932
1,2,3,4,7,8-Cl_6	6.553	2,3,7-Cl_3	7.149
1,2,3,7,8-Br_5	8.180	2,3,7,8-Br_4	8.824
1,2,3,7,8-Cl_5	7.102	2,3,7,8-Cl_4	8.000
1,2,4-Cl_3	4.886	2,7-Br_2	7.810
1,2,4,7,8-Br_5	7.770	2,8-Cl_2	5.495
1,2,4,7,8-Cl_5	5.959	3,7-Cl_2-2,8-Br_2	9.350
1,3,7,8-Br_4	8.699	3,7,8-Cl_3-2-Br	7.939
1,3,7,8-Cl_4	6.102	7,8-Cl_2-2,3-Br_2	8.830
1,3,7,8,9-Br_5	7.03		

Table 2. Binding Affinities of the Polychlorinated Dibenzofurans



molecule	pEC_{50}	molecule	pEC_{50}
1,2,3,4,7,8	6.638	2,3	5.326
1,2,3,4,8	6.921	2,3,4	4.721
1,2,3,6	6.456	2,3,4,6	6.456
1,2,3,6,7,8	6.569	2,3,4,6,7,8	7.328
1,2,3,7	6.959	2,3,4,7	7.602
1,2,3,7,8	7.128	2,3,4,7,8	7.824
1,2,3,7,9	6.398	2,3,4,7,9	6.699
1,2,4,6,7	7.169	2,3,4,8	6.699
1,2,4,6,7,8	5.081	2,3,6,8	6.658
1,2,4,6,8	5.509	2,3,7,8	7.387
1,2,4,7,8	5.886	2,3,8	6.000
1,2,4,7,9	4.699	2,6	3.609
1,2,4,8	5.000	2,6,7	6.347
1,3,4,7,8	6.699	2,8	3.590
1,3,6	5.357	3	4.377
1,3,6,8	6.658	4	3.000
1,3,8	4.071	no substituent	3.000
2	3.553		

Table 3. Binding Affinities of the Polychlorinated Biphenyls



molecule	pEC_{50}	molecule	pEC_{50}
2,2′,4,4′	3.886	2,3′,4,4′,5	5.041
2,2′,4,4′,5,5′	4.102	2,3′,4,4′,5,5′	4.796
2,3,3′,4,4′	5.367	2,3′,4,4′,5′,6	4.004
2,3,3′,4,4′,5	5.301	2′,3,4,4′,5	4.854
2,3,3′,4,4′,5′	5.149	3,3′,4,4′	6.149
2,3,4,4′,5	5.387	3,3′,4,4′,5	6.886
2,3,4,5	3.854	3,4,4′,5	4.553

Eliminating the descriptors of this correlated set (Table 5) is of interest; however, this highlights the difficulty of assigning relative importance to the individual correlated descriptors. Retaining only I_x^g, Q_{xx}, and Q_{zz} yields $r^2 = 0.759$ with $F = 73.2$, whereas retaining only m_0, Θ_1, Θ_2, and Θ_3 yields $r^2 = 0.746$ with $F = 50.6$. One expects that a molecular series of lower symmetry with hydrophobic substituent values that exhibited greater diversity would yield reduced correlation between the eigenvalues obtained about the property-field center and the second-order components calculated with respect to the principal geometric axes.

Centering the descriptor matrix about each of the column means and normalizing by the column standard deviations yields the principal components shown in Table 6. Retaining only the first four principal components accounting for 97% of the variance of the data yields $r^2 = 0.542$ with $F = 20.4$. Retaining the first two principal components as well as the fifth accounts for 84% of the data and yields $r^2 = 0.745$ with $F = 68.1$. Retaining only principal components accounting for the major variance of the data of a set of descriptors of mixed physical and chemical character normalized in this standard manner has been previously described as a strategy of questionable significance.[36] Finally, a partial least-squares (PLS) cross-validation leave-one-out calculation[37] yields $r^2 = 0.760$ with nine components. Of the three series, we have found, consistent with that previously observed,[29] correlation for the biphenyls to be poorly predictive. Eliminating the 14 biphenyl molecules from the cross-validated leave-one-out PLS calculation yields $r^2 = 0.772$ with eight components for the remaining 60 molecules. Since all 60 molecules are then planar, elimination of linearly dependent descriptors reduces the 11 to 8, and hence all descriptors are retained in the PLS calculation. Apparently

Table 4. Descriptor Matrix for Dibenzo-*p*-dioxins

molecule	I_x^g, Å^2	I_y^g, Å^2	I_z^g, Å^2	m_0	d, Å	Q_{xx}^c, Å^2	Q_{yy}^c, Å^2	Q_{zz}^c, Å^2	O_1, Å^2	O_2, Å^2	O_3, Å^2
1	50.14	187.68	237.82	3.36	0.76	19.55	38.35	57.90	5.18	11.50	16.67
2	55.80	198.78	254.58	5.14	1.81	31.46	77.83	109.29	6.12	11.85	17.97
3	65.17	210.01	275.18	7.52	0.00	50.87	122.87	173.74	6.76	16.34	23.10
4	56.70	204.45	261.15	5.74	1.07	33.19	97.26	130.45	5.63	15.96	21.59
5	57.69	210.01	267.69	6.33	0.45	35.36	115.96	151.32	5.59	18.11	23.70
6	55.17	213.61	268.79	7.21	0.52	34.75	150.20	184.94	4.61	20.78	25.39
7	53.92	209.97	263.89	5.74	0.44	27.38	112.58	139.95	4.62	19.60	24.21
8	54.82	193.28	248.09	4.55	1.36	29.28	60.12	89.40	6.24	11.57	17.81
9	58.44	207.15	265.59	7.21	0.22	43.39	127.95	171.33	5.88	17.84	23.72
10	56.70	204.45	261.15	5.74	0.19	33.19	97.55	130.75	5.69	17.08	22.76
11	54.11	207.10	261.20	6.32	0.48	32.18	122.35	154.52	5.04	19.18	24.22
12	52.99	204.41	257.40	5.14	0.40	25.62	93.85	119.46	4.94	18.13	23.08
13	58.39	207.14	265.53	7.27	0.93	43.07	128.02	171.08	5.51	17.16	22.67
14	47.47	194.18	241.65	3.65	1.41	14.57	62.32	76.89	3.60	15.49	19.09
15	52.02	198.80	250.81	4.55	1.20	23.45	74.95	98.40	4.69	15.52	20.20
16	52.85	204.52	257.36	5.14	0.70	24.79	94.41	119.20	4.13	18.57	22.70
17	49.68	207.14	256.82	5.43	0.93	20.38	117.34	137.72	3.51	20.99	24.50
18	49.21	204.44	253.66	4.55	0.75	17.69	90.49	108.18	3.73	19.51	23.24
19	50.84	213.59	264.43	6.32	0.01	23.62	144.87	168.49	3.74	22.93	26.66
20	50.22	209.96	260.17	5.14	0.01	19.85	109.04	128.89	3.86	21.21	25.07
21	48.43	200.84	249.26	4.54	0.01	16.60	91.05	107.66	3.00	20.72	23.72
22	48.30	198.83	247.13	3.95	0.53	15.82	71.91	87.73	3.72	18.21	21.92
23	50.52	211.80	262.32	5.73	0.18	21.72	126.96	148.69	3.76	22.16	25.92
24	50.36	210.90	261.26	5.44	0.31	20.77	118.01	138.78	3.80	21.64	25.44
25	50.52	211.80	262.31	5.73	0.56	21.73	126.91	148.64	3.79	21.84	25.63

Table 5. Correlation Matrix

	I_x	I_y	I_z	m_0	$\|d\|$	Q_{xx}^c	Q_{yy}^c	Q_{zz}^c	O_1	O_2	O_3
I_x	1	0.433	0.422	0.824	−0.176	0.948	0.538	0.651	0.672	0.049	0.153
I_y	0.433	1	0.938	0.442	−0.129	0.255	0.78	0.727	−0.046	0.77	0.75
I_z	0.422	0.938	1	0.375	−0.124	0.313	0.741	0.757	0.111	0.763	0.853
m_0	0.824	0.442	0.375	1	−0.338	0.828	0.796	0.833	0.288	0.332	0.323
$\|d\|$	−0.176	−0.129	−0.124	−0.338	1	−0.21	−0.336	−0.332	0.023	−0.377	−0.359
Q_{xx}^c	0.948	0.255	0.313	0.828	−0.21	1	0.488	0.645	0.746	−0.024	0.132
Q_{yy}^c	0.538	0.78	0.741	0.796	−0.336	0.488	1	0.969	−0.05	0.812	0.783
Q_{zz}^c	0.651	0.727	0.757	0.833	−0.332	0.645	0.969	1	0.162	0.722	0.774
O_1	0.672	−0.046	0.111	0.288	0.023	0.746	−0.05	0.162	1	−0.425	−0.143
O_2	0.049	0.77	0.763	0.332	−0.377	−0.024	0.812	0.722	−0.425	1	0.937
O_3	0.153	0.75	0.853	0.323	−0.359	0.132	0.783	0.774	−0.143	0.937	1

Table 6. Principal Components

	1	2	3	4	5	6	7	8	9	10	11
I_x	0.272	−0.419	0.062	−0.039	0.34	−0.468	0.567	−0.133	0.262	−0.043	0.003
I_y	0.338	0.171	0.319	0.051	0.61	0.287	−0.025	0.166	−0.349	−0.385	−0.022
I_z	0.342	0.149	0.376	0.301	0.136	−0.121	−0.463	−0.203	0.414	0.411	0.026
m_0	0.317	−0.259	−0.24	−0.444	0.029	0.21	−0.26	0.54	0.416	−0.019	−0.001
$\|d\|$	−0.153	−0.033	0.793	−0.493	−0.281	−0.074	0.072	0.116	−0.02	0.005	−0.002
Q^c_{xx}	0.249	−0.463	−0.037	−0.003	−0.125	−0.386	−0.411	0.038	−0.601	0.077	0.141
Q^c_{yy}	0.386	0.076	−0.081	−0.291	−0.113	0.387	0.25	−0.363	−0.147	0.306	0.531
Q^c_{zz}	0.394	−0.024	−0.043	−0.147	−0.321	0.106	−0.073	−0.471	0.02	−0.332	−0.607
O_1	0.068	−0.501	0.224	0.53	−0.261	0.483	0.222	0.195	−0.005	0.12	−0.101
O_2	0.306	0.383	−0.076	−0.002	−0.104	−0.175	0.31	0.399	−0.249	0.482	−0.404
O_3	0.325	0.292	0.031	0.281	−0.451	−0.247	0.096	0.236	0.14	−0.475	0.393
						variances					
	6.1124	2.8087	1.0688	0.6498	0.2858	0.0417	0.0202	0.0082	0.0041	0.0003	0.0001

the amount of information provided since the eight descriptors are not completely correlated contributes to an increase in the r^2. Eight components are not necessary to yield significant correlation. Five components will yield an $r^2 = 0.734$ for the 60 molecules.

In concluding this section it should be emphasized again that the primary intent of the present paper has been to detail certain of the formal relationships involving moment expansions as described in section 2. Since it has not focused on issues relating to the property molecular data sets and how such properties are inclusively described by the 11 moment descriptors presently proposed, this should be an interesting subject for future investigation.

4. DISCUSSION

The present paper has addressed some of the formal issues involved in obtaining three-dimensional moments of molecular property fields. Expansions that involve spatially distributed properties that sum to a finite value, i.e., with a nonvanishing zero-order moment, contrast with expansions for which this moment vanishes. For the former case, all moments above zero order are dependent upon the origin of expansion. As a consequence, attention has focused on the two centers of expansion previously identified, namely, the molecular centroid and what has been defined as the property-field center.[38] Moment expansions about the molecular centroid have been shown to yield first-order and second-order terms of the expansion that can be written in a form of the Eisenberg hydrophobic moment vector and the WHIM covariance matrix, respectively.

Molecular expansions about the property-field center yield a three-dimensional matrix descriptive of only the property-field distribution and not explicitly descriptive of the underlying molecular shape. This distinction is seen simply by the following hypothetical example. Assume that the property distribution remains constant as we change the underlying molecular structure, perhaps by displacing the atom positions slightly. For property-field values assigned to the atomic sites, this can arise simply if the only atoms displaced are those assigned vanishing property-field values. The second-order moments calculated about the property-field center and consequently the eigenvalues of the matrix composed of these components would remain unchanged. This is a simple consequence of the invariance of the property-field center to any changes in the underlying molecular structure that are not reflected in changes of the property-field distribution. Note the contrast between such behavior and the modifications introduced into the second-rank tensor or the WHIM covariance matrix calculated about the molecular centroid. Since the centroid is displaced, the components of this matrix will change even though the property-field distribution has remained unchanged. The WHIM descriptors, therefore, do involve an explicit description of the relationship between the property field and the underlying molecular structure.

Since expansion about the property-field center does not include such an explicit relationship, such a relationship has been introduced in a different manner. Additional descriptors involving second-order moments written in the frame of the principal geometric axes are introduced. Such moments include information that relates the property field to not only the location of the molecular centroid but also to the orientation of the molecular principal geometric axes. While it is certainly true that a modification in molecular structure will generally be accompanied by a modification of the property field, it is important to recognize that the two distributions are different if one is to develop systematic relationships between them.

Expansion about the property-field center provides the further capability of generating moment expansions to arbitrary order for the purposes of property-field moment comparisons. This is achieved by use of the property-field principal axes obtained in second order as a frame of reference for higher order. Recognition of chiral differences will, however, require a procedure that senses the principal axes.

Finally, the set of moment descriptors and related quantities proposed have been utilized in a QSAR with respect to the binding of 74 polyhalogenated aromatic molecules to the Ah cystolic receptor. The hydrophobic property-field moments provide descriptors that yield correlation as statistically significant as have been achieved previously.

The present paper has focused primarily on some of the formal issues involved in the utilization of molecular moments as molecular descriptors. The moments provide a simple three-dimensional description of fundamental molecular properties and can be used to expand scalar property-field distributions over a molecule into the realm of three dimensions. The discussion of section 2 has focused on two property fields, one of molecular density and the other of arbitrary character. It should be emphasized that the present development could be applied to any number of different

property fields simultaneously. For example, one might take grid-based electrostatic and steric values of CoMFA as two property fields and the grid-based hydrophobicity values of HINT as a third. Sums over the grid points would then yield moments with respect to the different distributions as well as moments and other quantities that related the distributions. While there would, of course, be a significant loss of detailed information, for a QSAR one would gain translation from a highly underdetermined statistical problem to one that was overdetermined with loss of the required molecular alignment step. Many options are, therefore, open with respect to the use of moments. Consequently, one might expect their continued use to be of value with regard to issues involving molecular comparison.

ACKNOWLEDGMENT

I thank Bruce Berne for directing me to the hydophobic moment literature of Eisenberg and for several useful discussions. I also acknowledge continuing discussions on these subjects with Mike Pitman.

REFERENCES AND NOTES

(1) Hansch, C.; Leo, A. *Exploring QSAR Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.

(2) *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, 1999.

(3) 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.

(4) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT: A new method of empirical hydrophobic field calculation for CoMFA. *J. Comput-Aided Mol. Des.* **1991**, *5*, 545−552.

(5) Baroni, M.; Constantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3DQSAR problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9−20.

(6) Kier, L. B.; Hall, L. H. In *Molecular Structure Description, The Electrotopological State*; Academic Press: San Diego, 1999; Chapter 4, p 62.

(7) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules *J. Med. Chem.* **1985**, *28*, 849−857.

(8) Todeschini, R.; Gramatica, P. New 3D Molecular Descriptors: The Whim theory and QSAR Applications. In *3D QSAR in Drug Design*; Kubinyi, H., Folkers, Martin, Y. C., Eds.; Kluwer/Escom: Dordrecht, 1998; Vol. 2, Part 3, p 355.

(9) Gancia, E.; Bravi, G.; Mascagni, P.; Zaliani, A. Global 3D-QSAR methods: MS−WHIM and autocorrelation. *J. Comput-Aided Mol. Des.* **2000**, *14*, 293−306.

(10) Pitman, M. C.; Huber, W. K.; Horn, H. H.; Rice, J. E.; Swope, W. C. FLASHFLOOD: A 3D Field-Based Similarity Search Method for Databases of Flexible Molecules. Submitted for publication.

(11) Turner, D. B.; Willett, P. Evaluation of the EVA descriptor for QSAR studies: 3. The use of a genetic algorithm to search for models with enhanced predictive properties (EVA_GA). *J. Comput-Aided Mol. Des.* **2000**, *14*, 1−21.

(12) Bursi, R.; Dao, T.; van Wijk, T.; de Gooyer, M.; Kellenbach, E.; Verwer, P. Comparative Spectra Analysis (CoSA): Spectra as Three-Dimensional Molecular Descriptors for the Prediction of Biological Activities. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 861−867.

(13) Good, A. C.; So, S.-S.; Richards, W. G. Structure−Activity Relationships from Molecular Similarity Matrixes. *J. Med. Chem.* **1993**, *36*, 433−438.

(14) Robert, D.; Amat, L.; Carbo-Dorca. Three-Dimensional Quantitative Structure−Activity Relationships from Tuned Molecular Quantum Similarity Measures: Prediction of the Corticosteroid-Binding Globulin Binding Affinity for a Steroid Family. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 333−334.

(15) Silverman, B. D.; Platt, D. E. Comparative Molecular Moment Analysis (CoMMA): 3D-QSAR without Molecular Superposition. *J. Med. Chem.* **1996**, *39*, 2129−2140.

(16) Further details of the procedure are provided on the Web site http://www.research.ibm.com/comma.

(17) Eisenberg, D.; Weiss, R. M.; Terwilleger, T. C. The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature* **1982**, *299*, 371−374.

(18) Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure−Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565−577.

(19) Herman, R. M. Center of Dispersion Force in HCL Interacting with Rare-Gas Atoms. *J. Chem. Phys*. **1966**, *44*, 1346−1352.

(20) Vegh, L. Scattering correlation in double ionization of helium by fast antiprotons and protons. *Phys. Rev. A* **1988**, *37*, 992−994.

(21) Schmelcher, P.; Cederbaum, L. S. Approximate constant of motion for molecular ions in a magnetic field. *Phys. Rev. A* **1989**, *40*, 3515−3523.

(22) Berant, Z.; Shahal, O.; Karpas, Z. Correlation between Measured and Calculated Mobilities of Ions: Sensitivity Analysis of the Fitting Procedure. *J. Phys. Chem.* **1991**, *95*, 7534−7538.

(23) Dhuicq, D.; Jugi, B.; Benoit, C.; Sidis, V. The He++H-2>HeH++H*-(2s, 2p) reaction: Study of the s and p contributions in a coincidence-correlation experiment. *J. Chem. Phys.* **1998**, *109*, 1998.

(24) Metral, C. J.; Janini, G. M.; Muschik, G. M.; Issaq, H. J. A Computer method for predicting the electrophoretic mobility of peptides. *HRC−J. High-Resolut. Chromatogr*. **1999**, *22*, 373−378.

(25) Taubin, G.; Bolle, R. M.; Cooper, D. B. Representing and Comparing Shapes Using Shape Polynomials. *IEEE Conference on Computer Vision and Pattern Recognition CVPR'89*; 1989; pp 510−516.

(26) Duda, R. O.; Hart, P. E. *Pattern Classification and Scene Analysis*; Wiley: New York, 1973; Chapter 9, p 364.

(27) Wagener, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of Molecular Surface Properties for Modeling *Corticosteroid Binding Globulin* and Cystolic *Ah* Receptor Activity by Neural Networks. *J. Am. Chem. Soc.* **1995**, *117*, 7769−7775.

(28) Clementi, S.; Cruciani, G.; Riganelli, D.; Valigi, R.; Constantino, G.; Baroni, M.; Wold, S. Autocorrelation as a tool for a congruent description of molecules in 3D-QSAR studies. *Pharm. Pharmacol. Lett*. **1993**, *3*, 5−8.

(29) Waller, C. L.; McKinney, J. D. Comparative Molecular Field Analysis of Polyhalogenated Dibenzo-p-dioxins, Dibenzofurans, and Biphenyls. *J. Med. Chem.* **1992**, *35*, 3660−3666.

(30) A detailed review of this work is presented in: Safe, S. Polychloriniated Biphenyls (PCBs), Dibenzo-p-Dioxins (PCDDs), and Related Compounds: Environmental and Mechanistic Considerations Which Support the Development of Toxic Equivalency Factors (TEFs). *Crit. Rev. Toxicol*. **1990**, *21*, 51−88.

(31) A brief historical review in a recent paper surveying molecular characterization can be found in: Livingstone, D. J. The Characterization of Chemical Structures Using Molecular Properties. A Survey. *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 195−209.

(32) Abraham, D. J. Kellog, G. E. *Hydrophobic Fields in 3D QSAR in Drug Design, Theory, Methods, and Applications*; Kubinyi, H., Ed.; Escom: Leiden, 1993; Part III. 3D QSAR Models, p 506.

(33) Clark, M.; Cramer, R. D. III; Van Opdenbosch, N. Validation of the Tripos 5.2 Force Field. *J. Comput. Chem*. **1989**, *10*, 982−1012.

(34) Matlab. The Mathworks Inc., 24 Prime Park Way, Natick, MA 01760.

(35) Matlab Statistics Toolbox, The Mathworks Inc., 24 Prime Park Way, Natick, MA 01760.

(36) Silverman, B. D. The Thirty-one Benchmark Steroids Revisited: Comparative Molecular Moment Analysis (CoMMA) with Principal Component Regression. *Quant. Struct.-Act. Relat.* **2000**, *19*, 237−246.

(37) SYBYL: Available from TRIPOS Inc., 1699 S. Hanley Road, St. Louis, MO 63144.

(38) The property-field center is actually a generalization of what has been called the "barycenter" (see, for example, http://spaceplace.jpl.nasa.gov/barycntr.htm).

CI000457S