Journal of Computer-Aided Molecular Design, 10 (1996) 513–520
ESCOM

J-CAMD 364

E-state fields: Applications to 3D QSAR

Glen E. Kellogg[a,*], Lemont B. Kier[a], Patrick Gaillard[b] and Lowell H. Hall[c]

[a]*Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, VA 23298-0540, U.S.A.*
[b]*School of Pharmacy, University of Lausanne, CH-1015 Lausanne, Switzerland*
[c]*Department of Chemistry, Eastern Nazarene College, Quincy, MA 02170, U.S.A.*

Received 30 January 1996
Accepted 15 June 1996

Keywords: CoMFA; Electrotopological state; QSAR

Summary

The derivation of a new 3D QSAR field based on the electrotopological state (E-state) formalism is described. A complementary index and its associated field, the HE-state, describing the polarity of hydrogens is also defined. These new fields are constructed from a nonempirical index that incorporates electronegativity, the inductive influence of neighboring atoms, and the topological state into a single atomistic descriptor. The classic CoMFA steroid test data set was examined with models incorporating the E-state and HE-state fields alone and in combination with steric, electrostatic and hydropathic fields. The single best model was the E-state/HE-state combination with $q^2 = 0.803$ (three components) and $r^2 = 0.979$. Using the E-state and/or HE-state fields with other fields consistently produced models with improved statistics, where the E-state fields provided a significant, if not dominant, contribution.

Introduction

Interaction between a drug molecule and a receptor, enzyme or binding site occurs as an event, governed by the collective effect of the electronic and topological structure. Atom-level descriptions of structure have traditionally been composed of two parts: a statement of electronic composition or probability and/or some quantitation of the steric environment or topology. It has been thought to be easier to consider these separately, to quantitate them and then to combine them in some fashion to reconstruct the atom structure. It would be advantageous if an atom-level description could embrace both these attributes. Indeed, the topology is a complex result of electronic structure. We can therefore think of a desirable description as one that unifies both these attributes. Such is the model of the 'atoms-in-molecule' structure that we have developed [1–3].

This unified description, called the electrotopological state of an atom or hydride group, is a nonempirical index encoding the electronegativity of the atom, the inductive influence of other atoms in the chemical graph, and its topological state. The derivation of this index, called the E-state index, is given in the Appendix. A number of reports [4–8] have revealed the value of this index in quantitative structure–activity studies.

It is our intent in the present study to expand upon the capabilities of this index by constructing an associated E-state field that purports to represent the spatial behavior of the E-state. This new field is then tested in the comparative molecular field analysis (CoMFA) paradigm of Cramer [9] to evaluate its information content and usability in three-dimensional QSAR. One of the advantages of field descriptors in QSAR, in contrast to atomistic descriptors, is that fields allow the consideration of more structurally diverse data sets. This is because the grid cage definition can be made invariant over all molecules of the data set (thus having an equivalent collection of descriptors), while for the atom-based indices the effects of structure variance must be inferred at atoms *common* to all molecules in the data set. We also describe for the first time the calculation of the hydrogen electrotopological state (HE-state) index and its associated field which complements the E-state by encoding electronic and topological information about the hydrogens. We show that the nonempirical E-state and HE-state fields are information rich and a potentially valuable addition to a large variety of 3D QSAR studies.

*To whom correspondence should be addressed.

Calculational details

Data set

The steroid data set of Cramer et al. [9], which includes 21 compounds, was used for this study. The corticosteroid-binding globulin (CBG) data were imported as the (dependent) biological data. Molecular structures and alignments were extracted from the SYBYL 6.2 release sample data and used without further modification.

E-state

The electrotopological state (E-state) was calculated for all non-hydrogen atoms in each molecule of the steroid learning set using algorithms [1–3] published previously and described in the Appendix to this work. These algorithms were used as incorporated in the Molconn-Z 3.0S program [10,11].

HE-state

The hydrogen electrotopological state (HE-state) was calculated for all heavy atoms that have bonded hydrogens using the algorithms described below in the Results section. To represent this property in three-dimensional space for contour grid maps, the HE-state was divided evenly among, and assigned to, the attached hydrogens.

E-state and HE-state maps

A three-dimensional grid cage was superimposed over each molecule in the study. Grid maps with both 1 and 2 Å spacing were explored. The value, E, at each grid point (t) is represented as

$$E_t = \Sigma\, S_i\, f(r_{it})$$

over all atoms (i) in the molecule, where S_i is the E-state (or HE-state) for atom i and $f(r_{it})$ is a function of the distance between the atom and the grid point t. A variety of functional forms for the distance behavior, f(r), were explored, including $1/r$, $1/r^2$, $1/r^3$, $1/r^4$, e^{-r}, and e^{-2r}. Since the $1/r^n$ functions are discontinuous at grid points close to atoms, field default values of zero were set for grid points within one van der Waals radius of any atom in the molecule.

TABLE 1
INTRINSIC STATE FOR HYDROGENS, I(H)

X-H	X ($\delta^v - \delta$)	I(H)
–OH	4	16.0
=NH	3	9.0
$-NH_2$	2	4.0
≡CH	2	4.0
–NH–	2	2.0
$=CH_2$	1	1.0
=CH–	1	1.0
–CH<	0	0.0
$-CH_2-$	0	0.0
$-CH_3$	0	0.0

3D QSAR

The three-dimensional QSAR was performed using the QSAR/CoMFA module of SYBYL 6.2 [12]. E-state and HE-state fields were imported into the SYBYL tables and examined with partial least squares (PLS) alone and in combination with each other, the HINT field [13,14] and the standard SYBYL steric/electrostatic fields. The HINT field was created with field default values of –15 for grid points in polar regions and +8 [14] for hydrophobic regions that were inside the van der Waals surface of the molecules. The analyses were performed with leave-one-out cross-validation and a column field filter of 0.5 kcal/mol.

Results

The E-state values of hydrogens (HE-state)

Since the E-state values are derived from a hydrogen-suppressed graph, the numerical values of reactive groups such as –OH subsume the contribution from the O and the H into a single value. Clearly, the H atom in this case is much more independent, reactive, and variable in electronic character than the H atom of an alkane. This great difference reveals itself most dramatically in situations where a hydrogen bond forms in either an inter- or intramolecular structure. In these situations, the strength of these bonds is largely a function of the structural influences on the hydrogen and oxygen atoms. Each however plays a different role and thus the group treatment of an OH fails to reveal the salient structural information. A significant advance in the E-state paradigm comes from the specific consideration of all the hydrogen atoms, with particular emphasis placed on the more polar hydrogens. We employ, in this report, the results of recent research into the meaningful parametrization of the hydrogen atom.

When we consider the intrinsic state of only the X-H fragment, the influence on -H is primarily derived from X. This influence is primarily from electronegativity, with only a small part coming from local topology. Based on the polarity of the hydrogen atoms, we suggest that the electronegativity effect should be accentuated. We propose the square, $(\delta^v - \delta)^2/\delta$, as the influence of X on H in X-H groups.

The intrinsic state would be calculated as $I(H) = (\delta^v - \delta)^2/\delta$. Table 1 summarizes some I(H) values. The E-state value for a hydrogen (HS_i) would then be calculated as

$$HS_i = I_i(H) + \Sigma\, \Delta I/r_{ij}^2$$

in a manner parallel to the scheme in the Appendix.

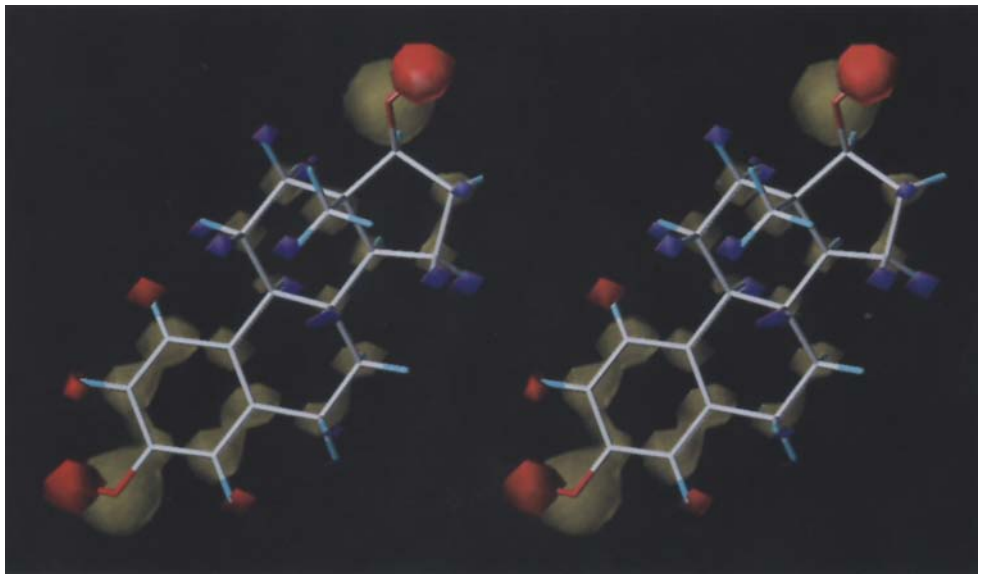

Fig. 1. Stereoview of the E-state (yellow, positive) and HE-state (red, positive and blue, negative) field contours for the estradiol molecule. The volumes of the enclosed contours indicate the relative magnitude of the associated field. A high positive E-state corresponds to a high electronegativity and reinforcing topology. A high positive HE-state signifies hydrogen polarity.

Representation of the E-state and the HE-state as 3D fields

While the electrotopological state (E-state) is an atomistic parameter set with significant applications for conventional QSAR, its utility has limitations. First, the effects of chemical substitutions and structure differences must be *inferred* from the examination of parameters on common atoms, rather than on the atoms of the substituents themselves. This, by necessity, requires fairly homogeneous sets of molecules in the QSAR analyses. Second, using atom-based parameters in QSAR is difficult to calculationally automate as molecular modeling programs often renumber and rename atoms during the model building process, and establishing the required one-to-one correspondence of atoms for QSAR manually is time-consuming. In contrast, the field technology invented by Cramer in CoMFA [9] circumvents these issues by describing each molecule in terms of a consistent set of grid points and inferred interaction potentials at the grid points.

Any atomistic property can be calculationally described as a field by superimposing a 3D cage of grid points over the molecule and calculating the 'field' effect on each of the grid points of all atoms in the molecule. Each of the grid points may be considered as a test atom or 'device' that measures the sum field effect at its point in space. The distance response functions of fields with known physical descriptions, e.g., electrostatic, are based on the associated interaction laws of the property. However, for less well-understood properties such as hydrophobicity [15] or more artificial properties such as the present E-state case, the distance response function is in dispute or unknown and best handled pragmatically. It may, in fact, be possible that the optimum distance response functions for these properties are data-set-dependent. In the next section, we examine a variety of functional forms for the E-state distance response in parallel CoMFA studies in an effort to establish the best representation.

The grid contour maps of the E-state fields for individual molecules are quite informative. In Fig. 1 we set out the E-state map and the HE-state map for estradiol contoured over a molecular model for estradiol. Yellow contours represent areas of 'high' positive E-state, red contours represent areas of 'high' positive HE-state, and light green contours represent areas of 'high' negative E-state. All properties are contoured at the same level. Thus, the relative volume of enclosed contours suggests the relative magnitudes of the E-state fields in these regions of space. It can be seen that the more polar substituent atoms to the steroid are coexistent with larger positive E-state and HE-state fields. A negative HE-state is localized on nonpolar hydrogens. Note that the aromatic hydrogens on the steroid A ring tend to be more polar than all the other hydrogens attached to carbons. This is consistent with the notion of aromatic hydrogens being more labile than aliphatic hydrogens.

Calibration of E-state and HE-state field distance functions

In order to understand the physical significance and statistical behavior of the E-state and HE-state fields, the functional form of the E-state distance behavior was explored in multiple runs at 1 and 2 Å grid spacing. The statistical results of these runs are summarized in Table 2.

Interestingly, the 2 Å q^2 results suggest that the optimum distance function for the E-state is 1/r, while the 1 Å results dispute this assertion with a consistent value of q^2 of 0.770–0.792 for $1/r^2$, $1/r^3$, $1/r^4$, e^{-r}, and e^{-2r}, with a slightly

TABLE 2
CALIBRATION OF E-STATE AND HE-STATE FIELD DISTANCE FUNCTIONS FOR THE STEROID DATA SET

Function form of E-state	E-state (1 Å) q^2 (components)	HE-state (1 Å) q^2 (components)	E-state (2 Å) q^2 (components)	HE-state (2 Å) q^2 (components)
$1/r$	0.726 (4)	0.685 (4)	0.792 (4)	0.705 (4)
$1/r^2$	0.770 (2)	0.665 (3)	0.743 (4)	0.717 (5)
$1/r^3$	0.791 (3)	0.661 (3)	0.668 (4)	0.636 (5)
$1/r^4$	0.774 (2)	0.599 (4)	0.590 (4)	0.366 (2)
e^{-r}	0.780 (2)	0.681 (3)	0.715 (3)	0.729 (3)
e^{-2r}	0.792 (2)	0.615 (3)	0.528 (2)	0.572 (3)

inferior q^2 of 0.726 for the 1/r function. For the HE-state field, the optimum distance function appears to be one of $1/r$, $1/r^2$, $1/r^3$, or e^{-r}. It should be pointed out that since the higher order functions decay more rapidly and will produce more noncontributing grid points with zero or near-zero values, the 2 Å models may be unfairly biased towards the lower order functions and will likely be more dependent on grid/molecule alignment. Thus, for the purpose of calibrating the E-state distance dependence functions, the 1 Å results are more reliable as they include more data in the PLS analyses. From these studies, it appears that $1/r^2$, $1/r^3$, or e^{-r} are essentially equivalent functions.

In CoMFA models using both the E-state and HE-state fields (Table 3), we describe the steroid data set in terms of these three functions. From the 1 Å models it appears that there is a very slight preference for the $1/r^3$ distance dependence as this is the model with the highest q^2 and this is the only model where adding the HE-state field to the E-state field increased q^2. The analogous 2 Å experiments indicate a very slight preference for either the $1/r^2$ or e^{-r} functions over the $1/r^3$ function, but most important is that the addition of the HE-state field to the E-state field fairly significantly improves q^2 for all the models, probably due to the increased number of correlation variables in the PLS. For all further experiments, the $1/r^3$ distance function will be used. The relative contributions of the E-state and the HE-state to these two-field models are also shown in Table 3. In general, the E-state and the HE-state contribute equally to two-field CoMFA models.

E-state/HE-state CoMFA models with other fields

The standard CoMFA analysis normally includes two fields: steric and electrostatic. The former is based on an adaptation of the Lennard-Jones potential function, while the latter is a simple electrostatic potential field calculated from the point charges of the atoms in the molecule. In addition, we have previously described [13,14,16] an empirical hydropathic field (HINT) based on the measured and calculated log P of molecules and fragments. Table 4 presents the results from multifield analyses of the steroid CoMFA data set.

The benchmark for CoMFA models is the standard CoMFA combination of steric and electrostatic fields in a single column. These results, under the conditions employed in this study, are reported in Table 4 for comparison purposes. Several of the new CoMFA models at both 1 and 2 Å grid spacing have better PLS statistics than the steric/electrostatic models. The differences are most striking in the 1 Å models, where (i) the E-state field alone produces a model with $q^2 = 0.791$; (ii) adding either the E-state or HE-state fields to the standard fields improves q^2 by 3–5% over standard CoMFA; and (iii) the combination of the E-state/HE-state fields yields a CoMFA model with $q^2 = 0.803$. Also of interest is that the E-state fields are 'greedy' in that they usually claim a 50% or more contribution when they are combined with other fields. In the case where the steric, electrostatic, E-state, and HE-state fields were included in a single model, the E-state fields have claimed a 60–70% contribution. This level of relative contribution indicates that the E-state and HE-state 3D fields are providing the bulk of the necessary information to describe the biological performance of the steroid data set. In fact, a review of Table 3 shows that all of the models constructed with only the E-state/HE-state fields are statistically superior to the standard CoMFA steric and electrostatic models.

The HINT (hydropathic) field was included in some of the above CoMFA models. Probably because conventional QSAR indicates a poor correlation of the CBG data with hydrophobicity parameters such as log P [9], the HINT field likewise does not add statistical information to the steroid CoMFA models. This field may, however, add useful chemical information for drug design for many other systems where hydrophobicity is a more important factor.

CoMFA coefficient field contours for the 1 Å E-state/HE-state model are presented in Fig. 2. The E-state/HE-state CoMFA coefficient field contours in Fig. 2 are a 3D predictive map of the effects of chemical modifications to the steroid molecules that will effect an increase or decrease in the biological activity of the steroid with respect to CGB. In this case, the mapped properties are the electrotopological state and/or the hydrogen electrotopological states of the proposed atoms. Cyan contours coincide with regions where an increasing (more positive) E-state will enhance the activity, while orange contours coincide with regions where an increasing E-state will decrease the activity. White contours coincide with regions where an

TABLE 3
CoMFA MODELS EMPLOYING BOTH E-STATE AND HE-STATE FIELDS

Function	1 Å model q^2 (components)	Contribution E-state/HE-state	2 Å model q^2 (components)	Contribution E-state/HE-state
$1/r^2$	0.767 (4)	0.489/0.511	0.779 (5)	0.386/0.614
$1/r^3$	0.803 (3)	0.522/0.478	0.763 (5)	0.616/0.384
e^{-r}	0.774 (3)	0.399/0.601	0.779 (3)	0.431/0.569

increasing (more positive) HE-state will enhance the activity, while magenta contours coincide with regions where an increasing HE-state will decrease the activity. The assignment of this map and its implications for ligand design will be described further in the Discussion section.

Discussion

The significance of the E-state values

The E-state index is a structure descriptor for an atom (or hydride group) within a covalently bonding molecule. It is derived from the counts of valence and σ bonding electrons in a hydrogen-suppressed chemical graph representing a molecule. The index is formulated to encode information about the electronegativity, π and lone-pair electron content, topological status and the environment of an atom within a molecule. The environment of an atom mitigates the electronic and topological structure of that atom in such a way that a characteristic state is produced.

If a structural change is introduced into a molecule, the environmental effect on a given atom produces some change in the electronic and topological state of that atom. The change in state is a function of the magnitude of the structure change and its proximity to the atom under scrutiny. Thus, in a series of molecules we can chronicle the influence of all structural changes on every atom in each molecule by using the E-state index.

The E-state values reflect the consequences of structure change at each atom. There arises then a mosaic of E-state values throughout the molecule in a given series. The expectation that discrete portions of a molecule are large-

TABLE 4
MULTIFIELD CoMFA MODELS

Model (fields)	1 Å models			2 Å models		
	q^2 (components)	r^2	Field contribution	q^2 (components)	r^2	Field contribution
Steric	0.736 (3)	0.950	0.472	0.733 (3)	0.944	0.505
Electrostatic			0.528			0.495
Hydropathic	0.664 (2)	0.918	1.000	0.622 (4)	0.968	1.000
E-state	0.791 (3)	0.964	1.000	0.668 (4)	0.961	1.000
HE-state	0.661 (3)	0.928	1.000	0.636 (5)	0.904	1.000
Steric	0.712 (3)	0.962	0.268	0.724 (4)	0.970	0.333
Electrostatic			0.353			0.380
Hydropathic			0.380			0.287
Steric	0.781 (2)	0.931	0.229	0.741 (4)	0.975	0.267
Electrostatic			0.224			0.303
E-state			0.547			0.430
Steric	0.762 (3)	0.973	0.202	0.801 (3)	0.960	0.295
Electrostatic			0.274			0.315
HE-state			0.668			0.390
Hydropathic	0.762 (2)	0.943	0.332	0.706 (3)	0.969	0.341
E-state			0.668			0.659
Hydropathic	0.697 (3)	0.968	0.382	0.728 (5)	0.976	0.525
HE-state			0.618			0.475
E-state	0.803 (3)	0.979	0.522	0.763 (5)	0.962	0.616
HE-state			0.478			0.384
Hydropathic	0.770 (2)	0.960	0.189	0.757 (4)	0.973	0.246
E-state			0.418			0.436
HE-state			0.478			0.318
Steric	0.791 (3)	0.976	0.121	0.779 (3)	0.966	0.198
Electrostatic			0.166			0.221
E-state			0.346			0.271
HE-state			0.367			0.309

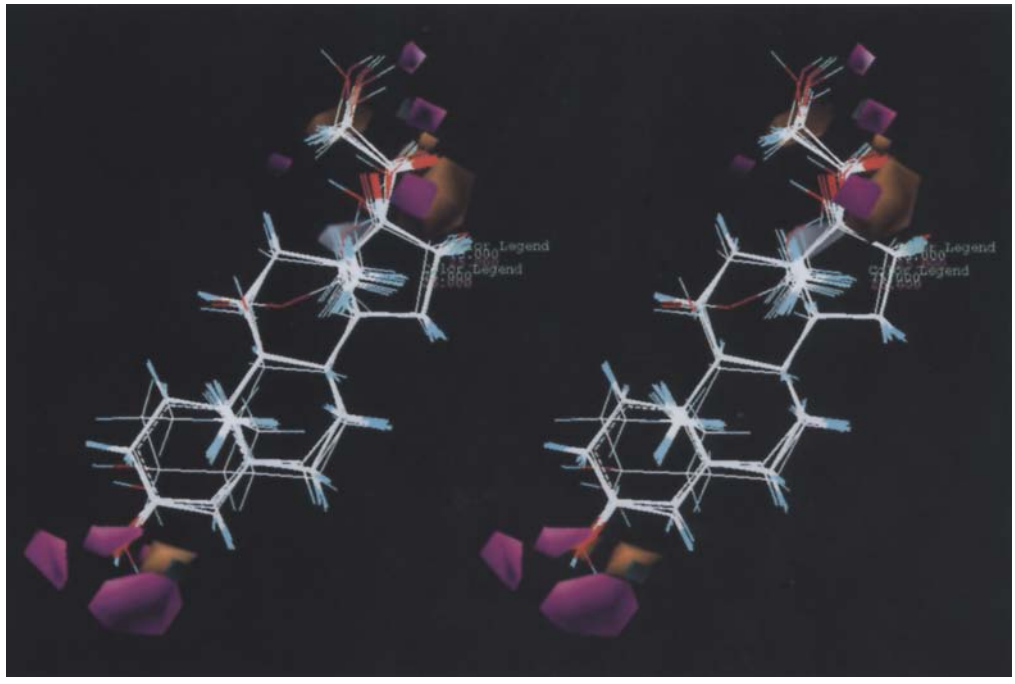

Fig. 2. Stereoview of the CoMFA coefficient contour maps for the E-state (orange and cyan plotted at 25% and 75% contribution) and the HE-state (magenta and white plotted at 25% and 75% contribution). The fields indicate regions where chemical modifications that increase the E-state or the HE-state can improve (cyan and white contour surfaces) the biological activity (binding affinity to CBGs).

ly responsible for receptor or enzyme binding can now be tested by studying this mosaic. With receptor binding or activity data in hand, comparisons may be made with changing E-state values throughout the molecule.

In general, we can say that with higher magnitudes of E-state values, we are describing atoms rich in π and lone-pair electrons and which are of mantle topology in the molecule. As a consequence, the interactions involving these atoms might be quite strong, i.e., electrostatic or hydrogen bonding. With E-state values at an intermediate range, we would look for dipolar forces to operate or we would expect that the atom is partially buried in the molecule and therefore less sterically accessible to interactions across space. Lower E-state values correspond to the dominant propensity for dispersion interactions. The regions of a molecule with E-state values in the 1.5–3.0 range may be considered to be possible lipophilic regions. At lower E-state values, there is a relative inertness to be expected as a consequence of π and lone-pair electron poverty or an appreciable buried atom status.

One test of these assignments is based on the projection of the E-state values into three-dimensional space around each molecule in the test series as was shown in Fig. 1 for estradiol.

An examination of the CoMFA coefficient contours in Fig. 2 in this context reveals a chemical-based assignment of the resulting field. In areas where the CoMFA coefficient contours indicate that the 'increased E-state' would enhance activity, two types of chemical modification can be employed. First, the substituents resident in the contours should, themselves, be replaced by more electronegative atoms or groups. Secondly, the neighboring atoms or groups can be adjusted to be more electron donating; e.g., adding a proximal methyl will enhance the electronegativity of neighboring polar groups. This latter effect underscores one of the more interesting aspects of the E-state formalism. Each atomic E-state value encodes the properties of its own atom as well as the topology of its environment. Because the electronic effects of substituents, fragments or groups on molecules are also coupled in much the same way, this feature of the E-state and the associated E-state fields is clearly useful for ligand design.

Information content of alternative fields

The typical CoMFA steric and electrostatic fields are surprisingly adept at modeling the biological activity of a drug molecule/receptor binding event. It is a commonly held notion that these are complex, nonlinear events comprised of a multitude of individual interactions between the atoms and/or groups of the drug, receptor site and solvent molecules. The steric and electrostatic fields are empirical, molecular-mechanics-based descriptors of the physics of interaction, i.e., London forces and Coulombic charges. In contrast, the electrotopological state is a nonempirical vehicle for atom electronegativity, electronic structure, and topology information. Thus, the E-state field encodes a somewhat different, more chemistry-based, information packet than either the electrostatic or steric fields, but does contain elements of both. This apparent overlap of information content is perhaps why

the E-state field captures equal fractions of field contribution from both steric and electrostatic fields when it is added as the third field to a CoMFA model. The three-field models for this steroid data set consistently show the E-state as the primary (ca. 50%) contributor.

The most significant difference between the E-state and HE-state fields is that the E-state is localized on and around heavy (non-hydrogen) atoms, while the HE-state is localized on and around the hydrogens. Thus, these two fields are particularly complementary. In this study, the hydrogen electrotopological state (HE-state) field appears to be less statistically significant than the E-state field in single-field models. This can be rationalized as follows: (i) the HE-state (by definition) ignores any atom, regardless of its physical, chemical or biological importance, that does not have hydrogens; and (ii) the HE-state is relevant in a smaller volume of space than the E-state and would be expected to contribute fewer nonzero parameters to the PLS model. For the same geometric reasons, the HE-state field probably will not normally be highly correlated with the E-state field. Note that the HE-state *does* improve cross-validation statistics when it is added to models using the standard (electrostatic and steric) field set, the HINT field, and/or the E-state field. Clearly, the HE-state encodes new, useful information that is missing from the other fields. Indeed, the best CoMFA models are generally those that include both the E-state and HE-state fields. For example, the best 1 Å CoMFA is a two-field model with the E-state and HE-state fields that has a q^2 of 0.803 and an r^2 of 0.979 with three components.

In general, we see the development of a new CoMFA paradigm where a variety of field types are explored in the process of creating a model. It appears that just as every receptor environment and/or biological activity measurement is different, each data set is also different. The availability of a collection of 3D QSAR fields, each tuned to a particular molecular or atomistic property that may be important in the binding process or the ultimate derived biological property, allows the creation of a more robust statistical model that is also more chemically and physically meaningful. The creation of the E-state fields contributes to this paradigm. We continue to explore the applicability of other fields in 3D QSAR.

Acknowledgements

The authors acknowledge the support of their institutions for the research. SYBYL was made available to Virginia Commonwealth University through a University Software grant from Tripos Inc.

References

1 Kier, L.B. and Hall, L.H., Pharm. Res., 7 (1990) 801.
2 Hall, L.H. and Kier, L.B., J. Chem. Inf. Comput. Sci., 31 (1991) 76.
3 Kier, L.B. and Hall, L.H., Adv. Drug Res., 22 (1992) 1.
4 Joshi, N. and Kier, L.B., Med. Chem. Res., 1 (1993) 409.
5 Tsantili-Kakoulidou, A., Kier, L.B. and Joshi, N., J. Chim. Phys., 89 (1992) 1729.
6 Tsantili-Kakoulidou, A. and Kier, L.B., Pharm. Res., 9 (1992) 1321.
7 Hall, L.H. and Kier, L.B., Med. Chem. Res., 2 (1992) 497.
8 Hall, L.H., Mohney, B.K. and Kier, L.B., Quant. Struct.-Act. Relatsh., 12 (1993) 44.
9 Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
10 Molconn-Z 3.0, Hall Associates Consulting, Quincy, MA, U.S.A.
11 Molconn-Z 3.0S, EduSoft, LC, Ashland, VA, U.S.A.
12 SYBYL 6.2, Tripos Inc., St. Louis, MO, U.S.A.
13 Kellogg, G.E., Semus, S.F. and Abraham, D.J., J. Comput.-Aided Mol. Design, 5 (1991) 545.
14 Nayak, V.R. and Kellogg, G.E., Med. Chem. Res., 3 (1994) 491.
15 Abraham, D.J. and Kellogg, G.E., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, The Netherlands, 1993, pp. 506–522.
16 Kellogg, G.E., Joshi, G.S. and Abraham, D.J., Med. Chem. Res., 1 (1992) 444.
17 Hinze, J. and Jaffe, H.H., J. Am. Chem. Soc., 84 (1962) 540.
18 Kier, L.B. and Hall, L.H., J. Pharm. Sci., 70 (1981) 583.

Appendix

Formulation of atom-centered indexes

Atom intrinsic value

We represent a molecule with a hydrogen-suppressed graph (i.e., a chemical graph) in which the atoms are identified as elements with certain valence states. The σ bonds are represented by dimensionless connections between atoms. The ingredients in the chemical graph are (i) the presence of an atom; (ii) its valence state, hence by inference the counts of σ, π and lone-pair electrons; and (iii) the degree of adjacency. It is from these ingredients that we create a parameter reflecting the electronic and topological state of an atom in the chemical graph. The topological state of an atom is a condition within a molecule which ranges between mantle atom and buried atom status. If we consider the case of neopentane, the four methyl groups are on the periphery or the mantle of the molecule. In contrast, the central atom is buried within the molecule. The quantitation of this attribute may be accomplished if we use the degree of adjacency, which is equal to the count of σ electrons contributed by an atom in the chemical graph, a value which we call δ in the

molecular connectivity paradigm. We can use the value δ directly, but it may be more convenient to reciprocate the δ to give a primary atom the largest value of this attribute. Thus, $1/\delta$ becomes an index of the topological state of an atom in a chemical graph. The electronic attribute of an atom in a chemical graph concerns the number of π, σ and lone-pair electrons associated with each atom. These are the reactive, interactive species which give rise to intermolecular events other than covalent bond information. The σ electrons, on the other hand, are far less reactive, forming the skeletal framework of the molecule, i.e., the topology.

In addition to this interaction, we are interested in the electronegativity of an atom, for this attribute is related to the interactions taking place between atoms within a molecule. In previous work on molecular connectivity, we have demonstrated that the Mulliken–Jaffe [17] valence state electronegativity could be closely correlated with the difference between the count of valence electrons and the count of σ electrons on an atom [18]. This is referred to as the Kier–Hall electronegativity. The valence electron count on an atom in a chemical graph is designated by δ in molecular connectivity formalism. It is equal to the count of valence electrons contributed by an atom in a molecule, Z^v, minus the count of hydrogens on that atom, H. Thus, $Z^v - H = \delta^v$. It follows that the Kier–Hall electronegativity is $\delta^v - \delta$. Dissecting each of these terms into their electronic composition, we obtain $\delta = \sigma$ electrons and $\delta^v = \sigma + \pi +$ lone-pair electrons. Therefore, $\delta^v - \delta = \pi +$ lone-pair electrons.

The expression for the Kier–Hall electronegativity is identical to the count of electrons recognized as being responsible for intermolecular events among organic molecules. To unify the electronic attribute $(\delta^v - \delta)$ and the topological attribute $(1/\delta)$, we may consider the product of these two to obtain a single index: $(\delta^v - \delta)/\delta$.

It is evident that in the case of the methyl, methylene, methine and quaternary fragments, the value $\delta^v - \delta$ is zero. This expression does not distinguish among the four adjacency patterns of these groups. The expression can be scaled to distinguish among these by adding 1 to the numerator. The expression becomes $(\delta^v - \delta + 1)/\delta$.

We can simplify this expression and, at the same time, scale it to values greater than 1 by adding 1 to the expression. This produces a value, I, for the intrinsic state of an atom in a chemical graph:

$$I = (\delta^v - \delta + 1)/\delta + 1 = (\delta^v + 1)/\delta \qquad \text{(A1)}$$

Influence on intrinsic state

The intrinsic state of an atom in a chemical graph reflects its electronic and topological attributes in the absence of interaction with the rest of the molecule. The influence of all the other atoms on this atom mitigates the value of this state. It is necessary to make some estimate of this perturbation in order to construct a model of the electronic and topological states of atoms in chemical graphs.

To estimate the perturbation of an intrinsic state value of an atom due to the influence of all other atoms, we dissect the graph into definable fragments. These are the sequences of skeletal bonds from atom i to atom j. Accordingly, we consider all paths emanating from a specific atom. These paths begin with an atom under study, i, and end with the last atom in the path, j. This constitutes a loge of two atoms separated by the remaining number of atoms in that path. The distance, r_{ij}, between the two extreme atoms is taken as the total number of atoms in the loge.

The perturbing influence of any atom, j, on atom i in any graph fragment may be estimated as the difference between the intrinsic states of the terminal atoms, $I_i - I_j$. If we assume that the electronic influence of atom j upon atom i in a loge follows an inverse-square law, we may quantitate the perturbation of atom i due to each loge to be $(I_i - I_j)/r_{ij}^2$. The total perturbation of I_i, expressed as ΔI_i, is therefore

$$\Delta I_i = [\Sigma(I_i - I_j)]/r_{ij}^2 \qquad \text{(A2)}$$

The summation is extended over the entire molecule. The state of each atom in a chemical graph due to the intrinsic state of that atom and the molecular field may be called the electrotopological state, E-state (S). It is calculated to be the sum of these two terms

$$S = \Delta I_i + I_i \qquad \text{(A3)}$$