communications in mathematical
match
and in computer chemistry

no. 44, October 2001
MATCDY (44) 215 - 235 (2001)
ISSN 0340 - 6253

DATABASE ORGANIZATION AND SEARCHING WITH E-STATE INDICES

Lemont B. Kier[a] and Lowell H. Hall[b]

[a]Department of Medicinal Chemistry, Virginia Commonwealth University
Richmond, Virginia, USA, phone 804 828 6451, e-mail Kier@hsc.vcu.edu

[b]Department of Chemistry, Eastern Nazarene College, Quincy,
Massachusetts, USA, phone 617 745 3550, e-mail Hall@enc.edu

Abstract. The electrotopological state (E-State) and its extension, the atom-type E-State, is presented as a representation of atom and molecular fragment structure useful for chemical database organization and management. An approach to database organization, using substituted esters and benzene derivatives as examples, reveals the descriptive power of the E-State paradigm. With a database, organized on the basis of structural relationships as described here, it is possible to search for similar molecular structures with potential for comparable activity. The searches using the atom-type E-State indices are demonstrated with several examples.

INTRODUCTION

Over the past decade there has been a significant growth in the use of similarity-based searching of databases in drug design [1-4]. The objective has been to organize a database of molecules according to some set of structure criteria so that compounds can be identified as being similar to a reference or target molecule. These similar compounds become candidates for screening or further analysis in the search process. The rationale is that compounds that are similar to a reference molecule are likely to be related to the behavior of the reference molecule in some sense. With the growth of combinatorial chemistry, the compounds in a database may be entirely or partially virtual, in other words they are synthesized *in silico*. As a result, there may be no property value information with the molecules, hence similarity is based entirely on

the structural descriptions chosen in a particular study. There is thus no way of evaluating similarity based on physical properties except by virtue of the future success of the drug design project employing this general method.

Lajiness has showed quite clearly that a random search through a list of molecules is inferior to a search through an organized database, based on its ability to generate similarity or diversity in a study [3]. Some form of encoding structure information should be present in order to meaningfully exploit a database. The code of structure information thus becomes the metric to evaluate similarity or its complement, diversity. This approach is not an exercise in multi-parameter QSAR modeling. With virtual molecules, many of the property values are unknown. The search is conducted by selecting a set of descriptors deemed important and finding the relative relation (for example, distance) of molecules to a reference molecule. The objective is to create a cluster of molecules of potential interest based on several sets of structure indices. Interesting compounds may appear that can be selected for screening or for further applications in the database search process.

The encoding and subsequent searching can be a browsing process, using E-State values or other information-rich indices, removing the need for carefully delineated structural features which are unknown or which can severely limit diversity. The choice of limiting distance values among molecules in the database makes it possible reduce the number of output molecules. A qualitative advantage of this process is the stimulation of the chemist's imagination [5].

There are a large number of descriptors that can be employed in the organization of a database. It is not our intention here to create a list of these or to make comparisons which may not be helpful, each method being suitable for different circumstances. Our intention however is to build on the use of the E-State atom-type index as a systematic organizer of a database, imparting a rich information source that can produce potentially useful structure patterns [6-8].

E-STATE DESCRIPTORS

An important objective of modeling is to obtain useful information about the structure features which influence the property being modeled. For this present case we use the molecular structure descriptors known as electrotopological state indices [7, 9-10]. The E-

State indices have been used to develop models for many activities and properties in both their atom-level and atom-type forms [7]. E-State QSAR models yield structure information which reveals structure features significantly related to activity. Further, the more recent development of hydrogen E-State values (and hydrogen atom-type E-State indices [7]) has extended the capability of the E-State as a powerful set of structure descriptors.

In this approach to structure representation, information is developed for each atom (such as >N- or =O or -Cl) and each hydride group (such as $-CH_3$, $-NH_2$, -OH) in the molecule. For simplicity both atoms and hydride groups are often called 'atoms'. The E-State index for atom i in a molecule is computed as follows:

$$S_i = I_i + \Sigma_j \Delta I_{ij}. \qquad \text{(sum over all other atoms j)} \qquad (1)$$

The perturbation term is as follows:

$$\Delta I_{ij} = (I_i - I_j)/r_{ij}^{\,2} \qquad (2)$$

in which r_{ij} is the number of atoms in the shortest path between atoms i and j. The E-State index is constituted from the atom intrinsic state (I_i) plus perturbations (ΔI_{ij}) by all other atoms in the molecule. In this manner each atom's E-State value contains electronic and topological structure information from all other atoms within the structure. The atoms closest to a given atom have the greatest influence on its E-State S value. Influence diminishes for atoms separated by a path of several bonds; the influence decreases as the square of the number of atoms in the path increases. A parallel development provides the basis for hydrogen atom-level E-State indices.

For a data set in which there is a common skeleton among the whole data set, the E-State values for these common skeletal atoms can be used directly as variables in seeking a QSAR model. The corresponding hydrogen atom-level E-State indices may be used in the same way.

THE ATOM-TYPE INDICES

An extension of the E-State method is the use of an atom-type index making it possible to study molecules of non-congeneric structure. Each atom is classified according to its valence state, the number of bonded hydrogens, and aromaticity [6]. For an atom-type index,

the E-State values are summed for all atoms of the same type in the molecule. The symbol for an atom-type index is $S^T(X)$ where X denotes the atom or hydride group. As examples we have S^T(-Cl) for a chloro, S^T(-OH) for a hydroxy, and S^T(...CH...) for an aromatic CH. The program Molconn-Z recognizes 80 atom-types [11]. Atom-type indices encode three distinct types of chemical structure information: 1) electron accessibility for the atom-type, 2) presence/absence of the atom-type and 3) count of atom-type present in the molecule. The atom-type E-State indices are used for heterogeneous data sets for structure-activity and for similarity analyses.

ORGANIZATION OF DATABASES

The atom type E-State values for an atom or groups in a molecule may be thought of as numerical components of basis vectors in a space or a manifold containing all possible atoms or groups. Each dimension is a parameter calculated for a particular atom or group. As an example, a set of molecules made up of alkanes, alcohols and glycols are defined by the E-State atom-types $-CH_3$, $-CH_2-$, and -OH shown in Table 1. The realm of this set of molecules can be defined by the three E-State atom-type indices corresponding to these groups shown in Figure 1. The alkanes, described by only two atom types, are located on one plane of the space. Likewise, the α,ω-glycols are found on a plane described by S^T(-OH) and $S^T(-CH_2)$. In contrast, the alcohols above methanol are found in the 3-D space defined by the above atom-type E-States plus $S^T(CH_3)$. In this space it is possible to find any of these three molecule types because of the organizing power of the atom-type E-State indices.

Within database subsets of molecules there are patterns of structure variation that are of interest in compound design. These patterns characterize the relative similarity or diversity within the subset. The atom-type indices make it possible to organize the subset in some manner which facilitates the design of other modifications, the selection of diverse structures for testing, cluster analysis based on structure and to conduct structure-activity analyses. A recent example using the polychlorobiphenyls showed how two atom-type indices could successfully organize the molecules in a modest size data set [12]. To further illustrate this organization ability and how the E-state indices can accomplish this, we examine two data sets in this study.

TABLE 1 Sum Atom Type E-State Values for a Set of Alcohols and Glycols

Molecule	S^T(-CH3) [a]	S^T(-CH2-) [b]	S^T(-OH) [c]
1. CH_3-OH	1.00	0.00	7.00
2. CH_3-CH_2-OH	1.68	-0.25	7.57
3. CH_3-CH_2-CH_2-OH	1.93	1.19	7.88
4. CH_3-CH_2-CH_2-CH_2-OH	2.05	2.38	8.07
5. HO-CH_2-CH_2-OH	0.00	-0.25	15.25
6. HO-CH_2-CH_2-CH_2-OH	0.00	0.69	15.81
7. CH_3-CH_2-CH_3	4.25	1.25	0.00
8. CH_3-CH_2-CH_2-CH_3	4.36	2.64	0.00

a. Symbol for the sum of E-State values for -CH_3 groups in the molecule.
b. Symbol for the sum of E-State values for -CH_2- groups in the molecule.
b. Symbol for the sum of E-State values for -OH groups in the molecule.

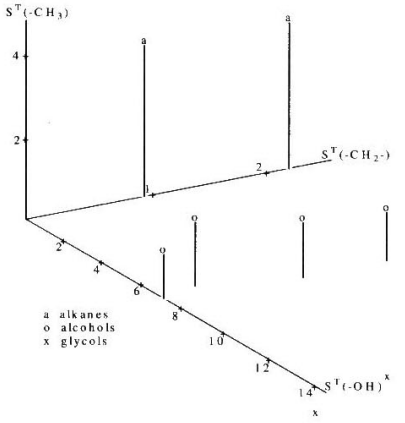


FIGURE 1. Three-dimensional realm of a set of alkanes, alkanols and glycols using atom-type E-State indices for all atoms present.

ESTER GROUP DERIVATIVES

Structural variations on the ester group often influence the ability to engage receptors or enzymes. The variations may be present on either the alcohol or the acid moiety, may be

near or distant from the ester fragment and may be of strong or weak electronic and steric influence. This matrix of variables presents an opportunity and a challenge to organize a virtual database of derivatives for the purpose of rational choices in compound design.

We have selected the atom-type E-State indices to illustrate the organizing power of the E-State structure representation. This set of esters is placed into the E-State space and projected onto the two-dimensional plane as shown in figure 2. Each atom-type E-State index represents the electron accessibility of that group in the molecule. S^T(>C=O) represents the electron accessibility at the carbonyl oxygen and S^T(-O-) represents the electron accessibility at the ether oxygen. Each substituent influences the electron accessibility at these oxygen sites. Methyl groups have low electronegativity and serve to increase the E-State value of both >C=O and -O- atoms. Hence the dimethyl derivative occupies the position of highest value for both indices, in the upper right-hand corner of the figure. On the other hand the difluoro derivatives act to decrease the electron accessibility at the oxygen atoms and, hence, occupies the lowest spot in the figure.

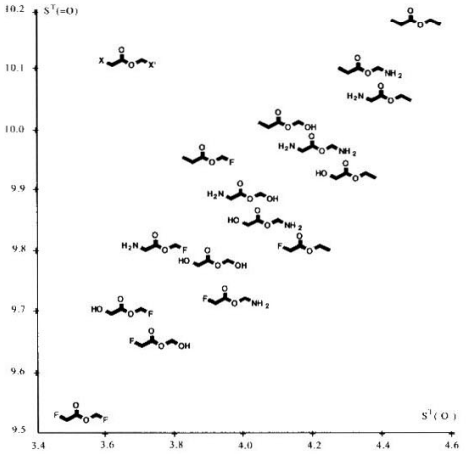


FIGURE 2. The two dimensional realm of substituted esters using atom-type E-State indices for the –O- and =O atoms as descriptors.

The electronic effects of the $-NH_2$ group are somewhat larger than that of $-CH_3$ but much less than -F. The two amino derivatives are located near the dimethyl derivative. To observe the proximity effect, consider esters of the CH_3COOCH_2X. For the series in which X varies from methyl to fluoro, the greater effect is exerted upon the ether oxygen. In the figure it is seen that the change in $S^T(-O-)$ is much greater than in $S^T(>C=O)$ for this series.

Relative placement of esters also makes good chemical sense. For example, $NH_2CH_2COOCH_2NH_2$ lies between $CH_3CH_2COOCH_2OH$ and $HOCH_2COOCH_2CH_3$. Likewise the dihydroxy compound, $HOCH_2COOCH_2OH$ lies between the amino, fluoro and the fluoro, amino ester. Each of the ester structures is located in a chemically meaningful place with respect to related structures. This organization is based on the atom-type E-State descriptors which represent electron accessibility. Because of this significant organization it would be possible to locate any ester based on the electron accessibility at the two oxygen atoms. Rational compound design is greatly facilitated by this kind of organizational ability.

Substituted Benzenes

A second set of compounds was analyzed, using the atom-type E-State indices. A series of substituted benzenes were described by the atom type indices for the aromatic CH and aromatic C atoms, Table 2 and Figure 3. We have selected molecules to portray the organization possible. The derivatives are organized according to the number of substituents, their relative electronegativity and the pattern that they occupy on the ring. Among the trisubstituted derivatives, the trifluoro- compounds have the lowest values of the two indices. A steady increase in these two values follows the decrease in the electronegativity of the substituents as illustrated by the location of the trihydroxy, the triamino and the trimethyl compounds. The values of the two indices exhibits a more dramatic increase when the pattern of variation is reduced to disubstitution. The monosubstituted derivatives have the highest values of the indices relative to their di- and tri- substituted analogs. From the table it is apparent that mixtures of substituents will fall into predicted positions in the plot. The organization using these two atom-type indices creates a pattern that is of potential value for any substituent hence is independent of the atom type. This makes it possible to search a database such as this for structurally similar (or diverse) molecules for alternatives with potentially comparable activity.

TABLE 2 Atom-Type E-State Values for Substituted Benzenes

No.	NAME	SaaCH	SAASC	No.	NAME	SaaCH	SAASC
1	Benzene	12.0000	0.0000	30	2-amino-3-fluorotoluene	4.7656	0.6968
2	Toluene	10.2616	1.3218	31	2-amino-4-fluorotoluene	4.3668	1.1431
3	Aniline	9.4871	0.8218	32	3-amino-5-fluoroytoluene	4.4647	1.0625
4	Phenol	8.7127	0.3218	33	2,6-dimethylphenol	5.7173	2.2940
5	Fluorobenzene	7.9382	-0.1782	34	2,4-dimethylphenol	5.5579	2.5008
6	o-Xylene	8.3565	2.7361	35	3,5-dimethylphenol	5.5136	2.5625
7	m-Xylene	8.4491	2.6759	36	2,6-difluorophenol	3.1540	-2.7894
8	p-Xylene	8.4815	2.6585	37	2,4-difluorophenol	2.5679	-2.1558
9	o-Aminotoluene	7.8043	2.0139	38	3,5-difluorophenol	2.3669	-1.9375
10	m-Aminotoluene	7.7996	2.0509	39	1-methyl-2-amino-3-hydroxy	5.1929	1.5440
11	p-Aminotoluene	7.7870	2.0785	40	1-methyl-2-amino-4-hydroxy	4.9362	1.8481
12	o-hydroxytoluene	7.2520	1.2917	41	1-methyl-2-amino-5-hydroxy	4.8912	1.8931
13	m-Hydroxytoluene	7.1502	1.4259	42	1-methyl-2-amino-6-hydroxy	5.0956	1.6412
14	p-Hydroxytoluene	7.0926	1.4985	43	1-methyl-3-amino-2-hydroxy	5.2901	1.4468
15	o-Fluorotoluene	6.6998	0.5694	44	1-methyl-3-amino-4-hydroxy	5.1306	1.6536
16	m-Fluorotoluene	6.5007	0.8009	45	1-methyl-3-amino-5-hydroxy	5.1306	1.6536
17	p-Fluorotoluene	6.3981	0.9185	46	1-methyl-3-amino-6-hydroxy	4.9884	1.7958
18	1,2,3-trimethylbenzene	6.3773	4.1829	47	1-methyl-4-amino-2-hydroxy	3.6618	-0.7008
19	1,2,4-trimethylbenzene	6.5023	4.1053	48	1-methyl-4-amino-3-hydrox	3.4673	-0.5064
20	1,3,5-trimetnylbenzene	6.5625	4.0625	49	1-fluoro-2-amino-3-hydroxy	3.9112	-0.9977
21	1,2,3-triaminobenzene	5.1929	1.5440	50	1-fluoro-2-amino-4-hydroxy	3.5195	-0.5586
22	1,2,4-triaminobenzene	5.0334	1.7508	51	1-fluoro-2-amino-5-hydroxy	3.6095	-0.6486
23	1,3,5-triaminobenzene	4.9892	1.8125	52	1-fluoro-2-amino-6-hydroxy	4.1056	-1.1921
24	1,2,3-trihydroxybenzene	4.0084	-1.0949	53	1-fluoro-3-amino-2-hydroxy	4.0084	-1.0949
25	1,2,4-trihydroxybenzene	3.5645	-0.6036	54	1-fluoro-3-amino-4-hydroxy	3.4223	-0.4614
26	1,3,5-trihydroxybenzene	3.4158	-0.4375	55	1-fluoro-3-amino-5-hydroxy	3.4223	-0.4614
27	1,2,3-trifluorobenzene	2.8240	-3.7338	56	1-fluoro-3-amino-6-hydroxy	3.7068	-0.7458
28	1,2,4-trifluorobenzene	2.0956	-2.9581	57	1-fluoro-4-amino-2-hydroxy	3.6618	-0.7008
29	1,3,5-trifluorobenzene	1.8425	-2.6875	58	1-fluoro-4-amino-3-hydroxy	3.4673	-0.5064

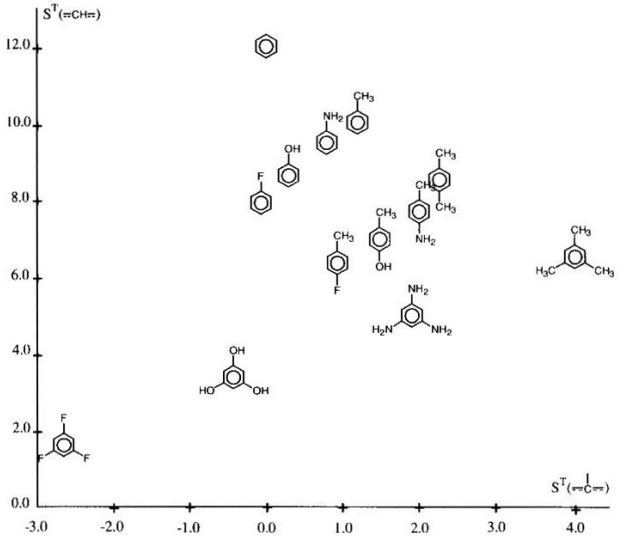


FIGURE 3. The two dimensional realm of substituted benzenes using the atom-type E-State indices for the substituted and unsubstituted aromatic carbon atoms as descriptors.

SIMILARITY SEARCHING USING THE ATOM-TYPE INDICES

A commonly held view is that molecules that are structurally similar, by some criterion, may have properties of comparable magnitude. In the context of biological properties, similarity presages comparable activities as ligands, substrates, inhibitors or other agents engaged in intermolecular encounters in living systems. In the area of drug design, this is the guiding principal in structure-activity modeling. A mathematical relationship is sought between a measured activity and a molecular property or structure representation. The purpose is to predict another structure, representing a candidate molecule, and to shed light on the structure-activity relationship. The rationale is that similarity may portend comparable behavior. The central element in this dialectic is similarity. Various investigators have pursued the similarity concept. Brown and Martin [13] have shown that the use of topological indices are often

superior to three-dimensional approaches. The E-State indices along with hydrogen E-State indices have been shown by Galvez to be rich in information for similarity selection among libraries of compounds[14].

The E-state concept is a prime candidate for the definition of similarity among molecules, molecular fragments and atoms-in-molecules[7]. One reason for this conjecture is that the E-state analysis produces numerical values which encode the extent of some attributes. The structure information may be represented at the atom level, the atom-type level and the bond type level, providing a broad basis for encoding molecular structure. We show here two examples of molecular similarity searches using two drug molecules as references and the atom-type E-State indices for the criterion of similarity. The database used is a modification of the Pomona MedChem database which contain 21,000 molecular structures.

Each atom-type E-State index was converted to a z score : $z_i = (x_i - \mu_i)/\sigma_i$, in which x_i is the i-th atom-type E-State in the database, μ_i is its mean and σ_i its standard deviation [4]. In this manner, each dimension in the structure space is put on a comparable basis of magnitude and spread. The calculations of the E-State indices was performed using Molconn-Z [11] and the searching was performed with the FORTRAN program ESEARCHZ developed by L. H. Hall. To assess structure similarity both the Euclidean distance (generalized Minkowski distance [4]) and the cosine formula are used [4].

$$d_{Euclidean} = \Sigma_j[(z_{ref} - z_j)^2]^{1/2} \quad (3)$$

$$cosine = \Sigma_j(z_{ref} * z_j)/[\Sigma_j(z_{ref})^2]^{1/2}[\Sigma_j(z_j)^2]^{1/2} \quad (4)$$

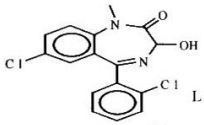
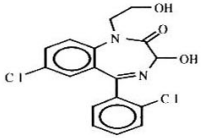
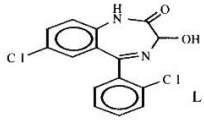
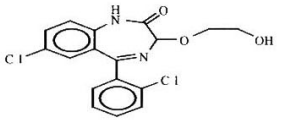
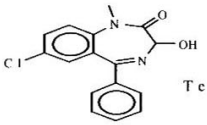
The sum is performed over the specified number of dimensions. Structures are ranked on their Euclidean distance. Both Euclidean distance and the cosine are reported in the following tables. The structures found closest to the target drug structure are listed in the two tables.

Lormetazepam

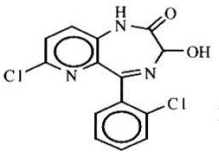
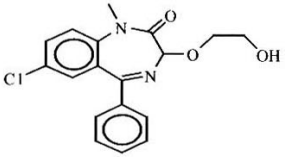
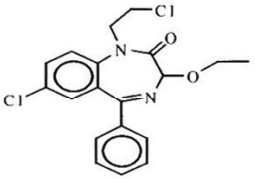
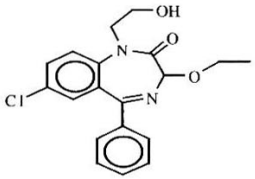
This drug is a sedative/ hypnotic of the benzodiazapine class. It is in a realm of structures in which there is much activity searching for analogues and bioisosteres. All ten E-State atom-types in this molecule were used in the search shown in Table 3 A. All of the close molecules found are similar to the reference molecule in respect to the presence of a seven-

membered ring and a phenyl substituent. A third common feature is the presence of two polar substituents on the diazapine ring. Those identified with a trade name have activities in the same category as the target molecule, Lormetazepam. Table 3 B shows the results of a search where the chlorine atom-type E-State was omitted in the structure description. This demonstrates the degree of selectivity possible by varying the pallet of atom types used in the search. Criteria of similarity are chosen by the investigator based on available information or intuition. By paring out descriptors from the pallet, it is possible to broaden the search and capture structures with greater diversity.

TABLE 3 A Database Similarity search with reference compound **Lormetazepam**, using all ten atoms types in the structure

Molecule Structure	Euclidean Distance (Arbitrary units)	Cosine
 Lormetazepam	reference	1.000
	1.26	0.956
 Lorazepam	1.26	0.948
	1.33	0.946
 Temazepam	1.36	0.940

Lormetazepam (continued)

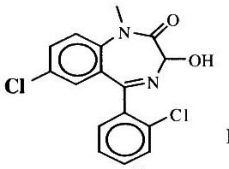
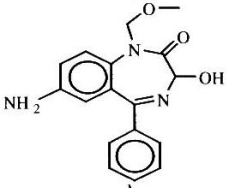
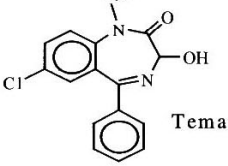
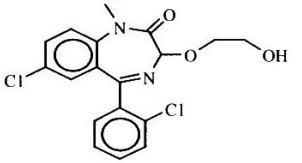
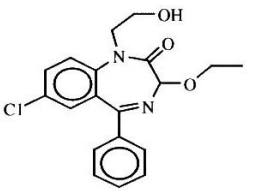
Molecule Structure	Euclidean Distance (Arbitrary units)	Cosine
 Lopirazepam	1.47	0.928
	1.50	0.924
	1.51	0.937
	1.52	0.923

Note: **Lormetazepam** found distance = 0.0

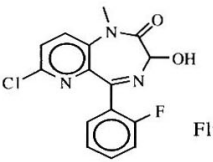
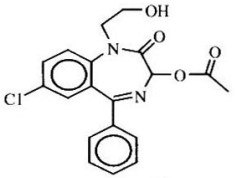
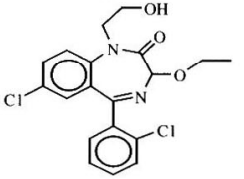
Next closest compound at distance = 1.72

Eleven more found up to distance = 1.85

Table 3 B Database Similarity search with reference compound **Lormetazepam**, using nine atoms types in the structure, leaving out –Cl

Molecule Structure	Euclidean Distance (Arbitrary units)	Cosine
 Lormetazepam	reference	1.000
	0.43	0.992
 Temazepam	0.45	0.991
	0.81	0.976
	0.85	0.974

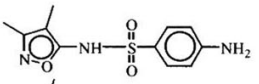
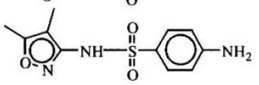
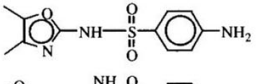
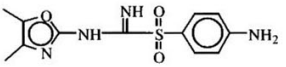
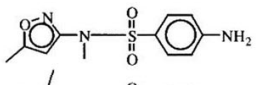
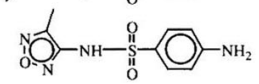
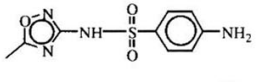
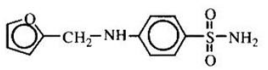
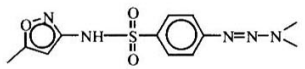
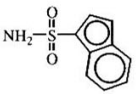
Lormetazepam (continued)

Molecule Structure	Euclidean Distance (Arbitrary units)	Cosine
 Flutemazepam	1.25	0.924
	1.25	0.934
	1.26	0.940

Sulfisoxazole

A second example of an effective similarity search using atom-type E-State indices is portrayed by the compound sulfisoxazole. This molecule is a prototype for antiinfectives of the sulfonamide class. These para-amino benzoic acid inhibitors have found their way into many different clinical classes, remote from their original clinical target. An effective search of a database is illustrated with this molecule using all of the atom-types found in the reference molecule. Table 4 reveals the results of a search through the database for the first nine compounds after the reference molecule. The principal variation at this level of search is on the sulfamido- aryl moiety, where the position of the heteroatoms varies in the series. The fragment separating the rings is also variable is the series. The molecules identified by a trade name are known to have activities comparable to the target molecule. The last compound found at this search level is significantly different in appearance but with a Euclidean distance

TABLE 4 Database Similarity search with reference compound Sulfisoxazole, using all nine atoms types in the structur

Molecule Structure	Euclidean Distance (Arbitrary units)	Cosine
	reference	1.000
	0.03	1.000
	0.33	0.999
	0.88	0.990
	1.26	0.984
	1.39	0.974
	1.49	0.969
	1.59	0.966
	2.04	0.946
	2.10	0.939

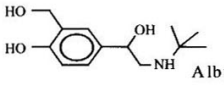
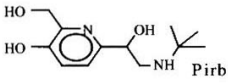
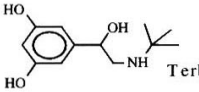
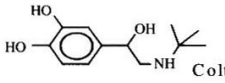
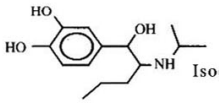
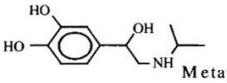
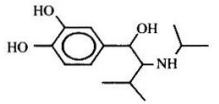
Next closest compound at distance = 2.17; ten more compounds found up to distance = 2.99 (cosine = 0.899). Sulfisoxazole found distance = 0.0

not that far from the next higher compound. This kind of deviation in the expected structures is often the opening to a new line of investigation in the drug design process. This is the ultimate goal of database searches and may be accomplished with this approach.

Albuterol

Albuterol is a beta-adrenergic, sympathomimetic agent used chiefly as a bronchodilator. This class of drugs finds wide use in the treatment of asthma. It is structurally related to norepinephrine whose action it mimics. A search through the database using all of the atom-type E-State indices in the molecule revealed a pattern of similar molecules. The principle variation in these structures is on the ring where hydroxyl and hydroxymethyl groups are found. Branching near the basic nitrogen occurs in some compounds among the more similar entries in the search. The compounds listed in Table 5 were all close in the ten-parameter space. Each molecule in the table identified by a trade name is known to be a bronchodilator operating by the same mechanism.

TABLE 5 Database Similarity search with reference compound Albuterol using all eight atoms types in the structure.

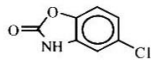
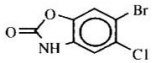
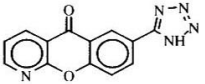
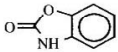
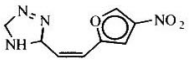
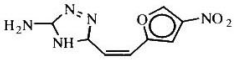
Molecule Structure	Euclidean Distance (Arbitrary units)	Cosine
 Albuterol	reference	1.000
 Pirbuterol	0.52	0.988
 Terbutaline	0.53	0.987
 Colterol	0.66	0.980
 Isoetharine	0.67	0.980
 Metaproterenol	0.95	0.957
	1.02	0.967

Chlorzoxazone

This drug is a representative of the centrally-acting skeletal muscle relaxants. An interesting display of structures is found to be close to the target molecule when all six atom-type E-State indices found in chlorzoxazone are used in the database search, Table 6 . The third molecule in the list, traxanox, is not a skeletal muscle relaxant but an antiasthmatic, antiallergenic agent. If all atom-type indices found in both this molecule and the target molecule were used in the search, perhaps this proximity in that space would not be as great. A

major change in the skeleton is found in this search producing the possibility of new lines of synthesis to achieve comparable activity. Also this suggests a way of uncovering secondary actions of molecules with similar structures by some criteria.

TABLE 6 Database similarity search with reference compound Chlorzoxazone using all six atoms types in the structure

Molecule Structure	Euclidean Distance (Arbitrary units)	Cosine
	reference	1.000
	0.52	0.996
	1.05	0.994
	1.27	0.973
	1.47	0.967
	1.49	0.965

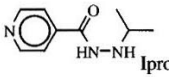
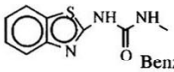
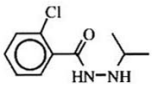
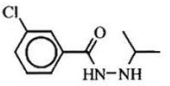
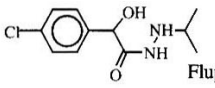
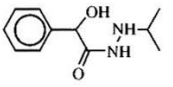
Next closest compound at Euclidean distance = 1.54
Chlorzoxazone found at distance = 0.0

Iproniazid

Iproniazid is an antidepressant acting by inhibiting monoamine oxidase which preserves serotonin, dopamine and norepinephrine, mediating the normal mental state of the patient. Compounds found near the target molecule have comparable activity while those further away have antimicrobial activity, Table 7 . This is understandable since the discovery of monoamine

oxidase inhibitors was the observation of side effects from antitubercular drugs. The variety found in the closest molecules to the reference molecule manifests itself in the side chain to the phenyl ring. The hydrazine is equivalent to a ureide in this search among the first two molecules. Similarly the hydroxyl and carbonyl groups present a similar presence in the series under these search criteria.

TABLE 7 . Database Similarity search with reference compound Iproniazid using all eight atoms types in the structure

Molecule Structure	Euclidean Distance (Arbitrary units)	Cosine
 Iproniazid	reference	1.000
 Benzthiazuron	0.93	0.934
	1.24	0.883
	1.26	0.878
 Flupirtine	1.28	0.875
	1.29	0.876

CONCLUSION

The electrotopological state (E-State) index is shown to contain information reflecting intermolecular accessibility of atoms and groups in a molecule. This information is encoded into a numerical value reflecting the electronegativity and the topology of structural fragments. The index for an atom is sensitive to the electronegativities of all other atoms in the molecule. This perturbing effect is carried through the network of atoms, described by a chemical graph.

The E-State indices may be used as individual atom-type descriptors thus making it possible to conduct structure-activity analyses with diverse structures. The E-State atom-type indices may also be used to organize a very diverse database of molecules into a coherent mosaic of molecules with a strong potential for navigating and searching in a logical way. This is demonstrated by the organizing of the substituted esters and benzenes into a database wherein any compound may be found by inspection. This capability forms the basis of the description of similarity by any criteria that are chosen. This is demonstrated by the search for similar compounds in a database, relative to a reference molecule. Thus lead compounds may serve as reference molecules in the search for potentially active congeners or bioisosteres, based upon an exploitable criterion of similarity. These are the opportunities made available using the E-State atom-type indices.

REFERENCES

[1] Willett, P. *Three-Dimensional Chemical Structure Handling*, John Wiley & Sons, New York 1991, 241p.

[2] Lajiness, M. S., Molecular Similarity-Based Methods for Selecting Compounds for Screening, in *Computational Chemical Graph Theory.* D. H. Rouvray, Ed., Nova Science, New York 1990, pp.300-312.

[3] Johnson. M. and Maggiora G. M., *Concepts and Applications of Molecular Similarity*, John Wiley & Sons, New York 1990, 278p.

[4] Willett, P., *Similarity and Clustering in Chemical Information Systems,* John Wiley & Sons, New York 1987, 315p.

[5] Warr, W., *Chemical Structures. The International Language of*

Chemistry, Springer, Berlin, 1988, 257p.

[6] Hall, L. H. and Kier, L. B. Electrotopological state indices for atom types: A Novel combination of electronic, topological and valence state information, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039-1045.

[7] Kier, L. B. and Hall, L. H. *Molecular Structure Description: The Electrotopological State*, Academic Press, San Diego, 1999, 145p.

[8] Hall, L. H. and Kier, L. B., Molecular Connectivity Indices for Database Analysis and Structure-Property Modeling, in *Topological Indices and Related Descriptors in QSAR and QSPR*, (J. Devillers and A.T. Balaban, Eds.) **1999**, 307-360.

[9] Kier, L. B. and Hall, L. H. An electrotopological state index for atoms in molecules, *Pharm. Res.* **1990,** *7*, 801-807

[10] Hall, L. H. and Kier, L. B. The electrotopological state: Structure information at the atomic level for molecular graphs, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76-83

[11] MOLCONN-Z may be obtained from Hall Associates Consulting, 2 Davis Street, Quincy, MA; SciVision Inc., 128 Spring Street, Lexington, MA 02173; Edusoft, LC, PO Box 1811, Ashland, VA 23005; and Tripos, Inc., 1699 South Hanley Rd, St. Louis, MO, 63144.

[12] Kier, L. B. and Hall, L. H. The E-state in database analysis: The PCB's as an example, *Il Farmaco*, **1999**, *54*, 346-551

[13] Brown, R. D. and Martin, Y. C. The use of structure-activity data to compare structure based clustering methods: Descriptors for use in compound selection, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572-584.

[14] Galvez, J., Garcia-Domenech, R., de Julian-Ortiz, J. V., and Soler,R., Topological approach to drug design, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272-284.