Prediction of Complement-Inhibitory Activity of Benzamidines Using Topological and Geometric Parameters

Subhash C. Basak,*,† Brian D. Gute,† and Shibnath Ghatak‡

Natural Resources Research Institute, The University of Minnesota, 5013 Miller Trunk Highway, Duluth, Minnesota 55811, and Department of Biology, Dana Laboratory, Tufts University, Medford, Massachusetts 02155

Received April 15, 1998

A hierarchical approach to quantitative structure−activity relationship (QSAR) modeling has been used to the estimate the complement-inhibitory potency of 105 benzamidines. This hierarchical approach uses topostructural, topochemical, and geometric parameters in a stepwise fashion to build increasingly more complex models. The results show that topostructural indices alone, specifically I^{D}, predict inhibitory potency reasonably well. The addition of topochemical and geometrical parameters to the set of descriptors provides only marginal improvement in predictive power. However, when taken alone, the geometric parameter ^{3D}W provides a more stable model than the topostructural one.

1. INTRODUCTION

A recent trend in structure−activity relationships (SAR) is the use of topological and geometric parameters in predicting the physicochemical, biochemical, and toxicological properties of molecules.[1−23] Topological indices (TIs) are numerical descriptors of molecular topology and encode information regarding the size, shape, branching, and symmetry of molecular graphs.[23] TIs and substructural parameters have been very useful in the development of quantitative structure−activity relationship (QSAR) models, in the quantification of the structural similarity of chemicals and in the similarity-based estimation of numerous physical and biological properties of diverse sets of molecules.[24−39] On the other hand, geometric variables such as total surface area, volume, and three-dimensional Wiener index have been employed in QSARs pertaining to biomedicinal and toxicological action of molecules with good results.[3,14,40−44]

One interesting area of research in biochemistry, pharmacology, and toxicology is the rationalization of the action of classes of chemicals with specialized modes of action. Specificity in enzymology, immunology, and toxicology arises out of specific structural features which lead to particular types of interactions between ligands and their biotargets. Topological and geometric parameters have been used in the development of QSARs of many groups of molecules with specific modes of action.[3,7,9,10,13,14,16,17,31−33,42−44]

Complement is a system of factors occurring in normal serum which are characteristically activated by antibody−antigen interactions and which subsequently mediate a number of biologically significant consequences.[45] The factors of the complement system include at least 20 chemically distinct serum proteins and glycoproteins. These factors, which normally exist in an inactive form, are activated by "classical" and "alternative" pathways. Both pathways generate macromolecular membrane attack complexes which lyse a variety of cells, bacteria, and viruses.[46] Products of this activation result in inflammatory reactions at the site of antibody−antigen interaction. This is especially pronounced in the case of organ specific and systemic autoimmune disorders. Therefore, control of unregulated complement activation is important, at least in the case of autoimmune disease.

Hansch and Yoshimoto[47] carried out a QSAR study of a set of 108 benzamidine derivatives using linear free-energy related (LFER) parameters. This series of compounds are inhibitors of the complement system. In view of the fact that LFER parameters are not routinely available for any arbitrary chemical, real or hypothetical, it was of interest to see whether computable parameters such as TIs and geometric indices can give a reasonable QSAR for the set of benzamidines. Therefore, in this paper we have carried out a comparative study of the utility of topological indices vis-à-vis calculated geometric parameters in predicting the complement-inhibitory potencies of this set of benzamidines.

Table 1. Conflicting Data for Structure and Log 1/*C* for Four Benzamidines

no.	X	obsd log 1/*C*
77	3-O$(CH_2)_3$OC_6H_4-3-NHCONHC$_6H_4$-3[a]-SO_2F	4.23
95	3-O$(CH_2)_3$OC_6H_4-3-NHCONHC$_6H_4$-3-SO_2F	4.51
97	3-O$(CH_2)_3$OC_6H_4-3-NHCOC$_6H_4$-4-SO_2F	4.57
108	3-O$(CH_2)_3$OC_6H_4-3-NHCOC$_6H_4$-4-SO_2F	5.21

[a] This SO_2F group should be *meta*- instead of *para*-.

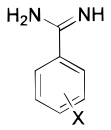

Figure 1. Neutral base structure for the 107 benzamidines.

* All correspondence to be addressed to Dr. Subhash C. Basak, Natural Resources Research Institute, University of Minnesota, Duluth, 5013 Miller Trunk Highway, Duluth, MN 55811. Phone: (218) 720-4230. E-mail: sbasak@wyle.nrri.umn.edu.

† The University of Minnesota.

‡ Tufts University.

10.1021/ci9800716 CCC: $18.00
Published on Web 12/04/1998

Table 2. Side-Chain Structures and Biological Property Data for 107 Benzamidines

		1/log *C*					1/log *C*		
no.	X	obsd	predict.[a]	resid	no.	X	obsd	predict.[a]	resid
1	3,5-$(OCH_3)_2$	−0.452	−0.367[b]	−0.085	55	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_3$-2-Cl-6-SO_2F	−0.255	−0.237	−0.018
2	2-CH_3	−0.444	−0.405	−0.040	56	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHC_6H_5$	−0.255	−0.249	−0.006
3	3,4-$(CH_3)_2$	−0.425	−0.389	−0.036	57	3-$O(CH_2)_2OC_6H_4$-2-$NHCONHC_6H_3$-2-Cl-5-SO_2F	−0.250	−0.236	−0.014
4	H	−0.418	−0.417	−0.002	58	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHCH_2C_6H_4$-4-SO_2F	−0.250	−0.228	−0.022
5	3-OH	−0.415	−0.402	−0.012	59	3-$O(CH_2)_3OC_6H_4$-2-NHCONH-C_6H_2-2,4-$(CH_3)_2$-5-SO_2F	−0.248	−0.229	−0.019
6	3-$NHCO(CH_2)_2C_6H_5$	−0.412	−0.302[b]	−0.110	60	3-$O(CH_2)_3OC_6H_4$-4-$COOCH_3$	−0.247	−0.271	0.025
7	3-CF_3	−0.410	−0.369	−0.041	61	3-$O(CH_2)_3OC_6H_3$-3-NO_2-4-CH_3	−0.245	−0.273	0.028
8	3-NO_2	−0.410	−0.378	−0.032	62	3-$O(CH_2)_3OC_6H_4$-3-CF_3	−0.245	−0.273	0.028
9	3-Br	−0.405	−0.401	−0.004	63	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHC_6H_4$-4-CH_3-3-SO_2F	−0.245	−0.229	−0.015
10	3-CH_3	−0.398	−0.402	0.004	64	3-$O(CH_2)_3OC_6H_4$-4-$NHCOC_6H_5$	−0.244	−0.246	0.002
11	3-OCH_3	−0.397	−0.389	−0.008	65	3-$O(CH_2)_3OC_6H_4$-2-$NHCOCH_2OC_6H_4$-4-SO_2F	−0.244	−0.227	−0.017
12	3-$CH_2C_6H_5$	−0.373	−0.339	−0.034	66	3-$O(CH_2)_3OC_6H_4$-4-$NHCOC_6H_4$-4-OCH_3	−0.243	−0.236	−0.007
13	3,5-$(CH_3)_2$	−0.361	−0.389	0.028	67	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_4$-3-SO_2F	−0.243	−0.238	−0.005
14	3-OC_3H_7	−0.355	−0.362	0.007	68	3-$O(CH_2)_3OC_6H_4$-2-$NHCOCH_2C_6H_4$-4-SO_2F	−0.243	−0.233	−0.010
15	3-*i*-C_5H_{11}	−0.355	−0.353	−0.002	69	3-$O(CH_2)_3OC_6H_4$-3-$COOCH_3$	−0.242	−0.272	0.030
16	3-OC_4H_9	−0.351	−0.349	−0.001	70	3-$O(CH_2)_3OC_6H_4$-2-$NHCO(CH_2)_2C_6H_4$-4-SO_2F	−0.242	−0.227	−0.014
17	3-C_4H_9	−0.338	−0.362	0.024	71	3-$O(CH_2)_3OC_6H_4$-4-$NHCOC_6H_4$-4-NO_2	−0.239	−0.232	−0.007
18	3-CH=CHC_6H_5	−0.339	−0.325	−0.014	72	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_4$-4-NO_2	−0.239	−0.241	0.002
19	3-$OCH_2C_6H_5$	−0.331	−0.326	−0.005	73	3-$O(CH_2)_3OC_6H_4$-4-$NHCONHC_6H_5$	−0.237	−0.241	0.004
20	3-$(CH_2)_2C_6H_5$	−0.330	−0.326	−0.004	74	3-$O(CH_2)_3OC_6H_4$-4-$NHCOC_6H_4$-3-NO_2	−0.237	−0.233	−0.005
21	3-OC_6H_{13}	−0.329	−0.327	−0.002	75	3-$O(CH_2)_3OC_6H_4$-2-$NHCO(CH_2)_4C_6H_4$-4-SO_2F	−0.237	−0.217	−0.020
22	3-$O(CH_2)_4OC_6H_5$	−0.325	−0.288	−0.037	76	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHC_6H_4$-4-SO_2F	−0.237	−0.233	−0.004
23	3-$O(CH_2)_2OC_6H_5$	−0.323	−0.306	−0.017	77	3-$O(CH_2)_3OC_6H_4$-3-$NHCONHC_6H_4$-4-SO_2F	−0.236	−0.225	−0.011
24	3-C_6H_5	−0.323	−0.347	0.025	78	3-$O(CH_2)_3OC_6H_4$-2-$NHCONH(CH_2)_2C_6H_4$-4-SO_2F	−0.236	−0.223	−0.014
25	3-$O(CH_2)_3OC_6H_4$-4-COOH	−0.321	−0.277	−0.044	79	3-$O(CH_2)_4OC_6H_4$-3-$NHCOC_6H_4$-4-SO_2F	−0.236	−0.223	−0.013
26	3-OC_5H_{11}	−0.320	−0.338	0.017	80	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHC_6H_3$-4-Cl-3-SO_2F	−0.235	−0.229	−0.006
27	3-O-*i*-C_5H_{11}	−0.318	−0.341	0.022	81	3-$O(CH_2)_4OC_6H_4$-2-$NHCOC_6H_3$-4-CH_3-3-SO_2F	−0.235	−0.229	−0.006
28	3-$O(CH_2)_2OC_{10}H_7$-α	−0.312	−0.283	−0.030	82	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_2$-2,4-$(CH_3)_2$-5-SO_2F	−0.234	−0.233	−0.001
29	3-$O(CH_2)_4OC_6H_4$-4-NH_2	−0.306	−0.282	−0.024	83	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_2$-2,4-Cl_2-5-SO_2F	−0.234	−0.233	−0.001
30	3-$(CH_2)_4C_6H_5$	−0.302	−0.306	0.004	84	3-$(CH_2)_4C_6H_4$-2-$NHCONHC_6H_4$-3-SO_2F	−0.234	−0.239	0.005
31	3-$O(CH_2)_3OC_6H_4$-4-NO_2	−0.301	−0.277	−0.024	85	3-$O(CH_2)_3OC_6H_4$-3-$NHCONHC_6H_4$-4-OCH_3	−0.233	−0.237	0.004
32	3-$O(CH_2)_3OC_6H_4$-4-NH_2	−0.300	−0.290	−0.010	86	3-$(CH_2)_4C_6H_4$-2-$NHCONHC_6H_4$-4-SO_2F	−0.233	−0.239	0.007
33	3-$(CH_2)_2$-4-C_5H_4N	−0.299	−0.326	0.026	87	3-$O(CH_2)_3OC_6H_4$-4-$NHCOC_6H_4$-4-Cl	−0.232	−0.241	0.009
34	3-$O(CH_2)_3OC_6H_5$	−0.299	−0.297	−0.003	88	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_3$-2-CH_3-5-SO_2F	−0.232	−0.236	0.004
35	3-$O(CH_2)_3C_6H_5$	−0.296	−0.306	0.010	89	3-$O(CH_2)_4OC_6H_4$-4-$NHCONHC_6H_3$-2-OCH_3-5-SO_2F	−0.232	−0.214	−0.018
36	3-$(CH_2)_2$-3-C_5H_4N	−0.294	−0.326	0.032	90	3-$O(CH_2)_3OC_6H_4$-4-C_6H_5	−0.230	−0.261	0.031
37	3-$(CH_2)_4C_6H_4$-4-NHAc	−0.294	−0.273	−0.021	91	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHC_6H_4$-3-SO_2F	−0.230	−0.233	0.003
38	3-$(CH_2)_2$-2-C_5H_4N	−0.291	−0.326	0.035	92	3-$O(CH_2)_3OC_6H_4$-3-$NHCOC_6H_4$-3-SO_2F	−0.230	−0.230	−0.000
39	3-$O(CH_2)_3OC_6H_4$-2-NH_2	−0.283	−0.291	0.009	93	3-$O(CH_2)_2OC_6H_4$-3-$NHCOC_6H_4$-3-SO_2F	−0.229	−0.236	0.007
40	3-$O(CH_2)_3OC_6H_4$-4-NHAc	−0.278	−0.265	−0.012	94	3-$O(CH_2)_3OC_6H_4$-4-CH_3-3-$NHCOC_6H_4$-4-SO_2F	−0.229	−0.226	−0.003
41	3-$(CH_2)_4$-3-C_5H_4N	−0.276	−0.306	0.030	95	3-$O(CH_2)_3OC_6H_4$-3-$NHCONHC_6H_4$-3-SO_2F	−0.222	−0.226	0.004
42	3-$O(CH_2)_4C_6H_5$	−0.276	−0.297	0.020	96	3-$O(CH_2)_3OC_6H_4$-3-$NHCOCH_2C_6H_4$-4-SO_2F	−0.220	−0.226	0.006
43	3-$O(CH_2)_3OC_6H_4$-3-NHAc	−0.270	−0.267	−0.003	97	3-$O(CH_2)_3OC_6H_4$-3-$NHCOC_6H_4$-4-SO_2F	−0.219	−0.229	0.010
44	3-$O(CH_2)_3OC_6H_3$-3,4-Cl_2	−0.265	−0.283	0.018	98	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHC_6H_3$-2-Cl-5-SO_2F	−0.217	−0.230	0.013
45	3-$O(CH_2)_3OC_6H_4$-3-NH_2	−0.265	−0.290	0.025	99	3-$O(CH_2)_3OC_6H_4$-3-$NHCOCH_2OC_6H_4$-4-SO_2F	−0.217	−0.219	0.002
46	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_4$-4-SO_2F	−0.265	−0.237	−0.028	100	3-$O(CH_2)_2OC_6H_4$-3-$NHCONHC_6H_4$-4-SO_2F	−0.216	−0.231	0.015
47	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_5$	−0.265	−0.253	−0.012	101	3-$O(CH_2)_4OC_6H_4$-3-$NHCONHC_6H_4$-4-SO_2F	−0.215	−0.220	0.005
48	3-$O(CH_2)_3OC_6H_4$-4-OCH_3	−0.262	−0.283	0.022	102	3-$O(CH_2)_3OC_6H_4$-3-$NHCOC_6H_4$-4-NO_2	−0.214	−0.233	0.019
49	3-$O(CH_2)_4OC_6H_4$-4-$NHCONHC_6H_4$-4-SO_2F	−0.260	−0.219	−0.040	103	3-$O(CH_2)_2OC_6H_4$-3-$NHCOC_6H_4$-4-SO_2F	−0.214	−0.235	0.021
50	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_3$-2-OCH_3-5-SO_2F	−0.260	−0.233	−0.027	104	3-$O(CH_2)_4OC_6H_4$-2-$NHCONHC_6H_3$-2-Cl-5-SO_2F	−0.207	−0.225	0.018
51	3-$O(CH_2)_3OC_6H_4$-4-Cl	−0.257	−0.290	0.033	105	3-$O(CH_2)_3OC_6H_4$-3-$NHCONHC_6H_4$-4-NO_2	−0.204	−0.230	0.025
52	3-$O(CH_2)_3OC_6H_4$-2-NO_2	−0.257	−0.281	0.024	106	3-$O(CH_2)_3OC_6H_4$-4-CH_3-3-$NHCONHC_6H_4$-4-SO_2F	−0.204	−0.223	0.018
53	3-$O(CH_2)_3OC_6H_4$-3-NO_2	−0.257	−0.278	0.021	107	3-$O(CH_2)_3OC_6H_4$-3-$NHCONH(CH_2)_2C_6H_4$-4-SO_2F	−0.193	−0.215	0.022
54	3-$O(CH_2)_3OC_6H_4$-3-OCH_3	−0.256	−0.283	0.027					

[a] Predicted values based on eq 2. [b] Values for compounds excluded from final modeling, provided to show lack of fit.

2. METHODS

2.1. Database. The 107 benzamidines used in this study are those presented in the work of Hansch and Yoshimoto.[47] These data were compiled from a series of five articles by B. R. Baker,[48−52] in which Baker and his students determined experimentally the inhibition of guinea pig complement by benzamidines. Hansch and Yoshimoto provide the structures and measured log 1/*C* values, where *C* is the micromolar concentration for 50% inhibition of complement (I_{50}), for 108 benzamidines. The numbered ordering used by Hansch and Yoshimoto will be used in this manuscript as well for ease of comparison. In the process of coding the data, it became evident that two of the compounds had structural duplicates with distinctly different values for log 1/*C* (see Table 1). Through close examination of Baker's work, it became evident that there was a typographic mistake in compound 77, while the error in compound 108 could not be accounted for. Thus, compound 108 was discarded from the set, leaving 107 benzamidine derivatives. The base structure of the benzamidines is presented in Figure 1, while their side chains and biological activities, both measured and estimated, are presented in Table 2.

Table 3. Symbols and Definitions of Topological and Geometrical Parameters

I_D^W	information index for the magnitudes of distances between all possible pairs of vertices of a graph
$\bar{I}_D^W$	mean information index for the magnitude of distance
W	Wiener index = half-sum of the off-diagonal elements of the distance matrix of a graph
I^D	degree complexity
H^V	graph vertex complexity
H^D	graph distance complexity
$\overline{IC}$	information content of the distance matrix partitioned by frequency of occurrences of distance h
I_{ORB}	information content or complexity of the hydrogen-suppressed graph at its maximum neighborhood of vertices
O	order of neighborhood when IC_r reaches its maximum value for the hydrogen-filled graph
M_1	A Zagreb group parameter = sum of square of degree over all vertices
M_2	A Zagreb group parameter = sum of cross-product of degrees over all neighboring (connected) vertices
IC_r	mean information content or complexity of a graph based on the rth ($r = 0-6$) order neighborhood of vertices in a hydrogen-filled graph
SIC_r	structural information content for rth ($r = 0-6$) order neighborhood of vertices in a hydrogen-filled graph
CIC_r	complementary information content for rth ($r = 0-6$) order neighborhood of vertices in a hydrogen-filled graph
${}^h\chi$	path connectivity index of order $h = 0-6$
${}^h\chi_C$	cluster connectivity index of order $h = 3-6$
${}^h\chi_{PC}$	path-cluster connectivity index of order $h = 4-6$
${}^h\chi_{Ch}$	chain connectivity index of order $h = 6$
${}^h\chi^b$	bond path connectivity index of order $h = 0-6$
${}^h\chi_C^b$	bond cluster connectivity index of order $h = 3-6$
${}^h\chi_{Ch}^b$	bond chain connectivity index of order $h = 6$
${}^h\chi_{PC}^b$	bond path-cluster connectivity index of order $h = 4-6$
${}^h\chi^v$	valence path connectivity index of order $h = 0-6$
${}^h\chi_C^v$	valence cluster connectivity index of order $h = 3-6$
${}^h\chi_{Ch}^v$	valence chain connectivity index of order $h = 6$
${}^h\chi_{PC}^v$	valence path-cluster connectivity index of order $h = 4-6$
P_h	number of paths of length $h = 0-10$
J	Balaban's J index based on distance
J^B	Balaban's J index based on bond types
J^X	Balaban's J index based on relative electronegativities
J^Y	Balaban's J index based on relative covalent radii
V_W	van der Waal's volume
${}^{3D}W$	3-D Wiener number for the hydrogen-suppressed geometric distance matrix
${}^{3D}W_H$	3-D Wiener number for the hydrogen-filled geometric distance matrix

2.2. Calculation of Topological Indices (TIs). Topological indices used in this study have been calculated by POLLY 2.3.[53] These indices include Wiener index,[54] connectivity indices,[16,55] information theoretic indices defined on distance matrices of graphs,[56,57] and a set of parameters derived on the neighborhood complexity of vertices in hydrogen-filled molecular graphs[10,58−60] as well as Balaban's J indices.[61−63] Table 3 gives brief definitions for the topological indices included in this study.

2.3. Calculation of Geometrical Indices. Volume (V_w) was calculated using the *Sybyl*[64] package from Tripos Associates, Inc. The 3-D Wiener numbers were calculated using *Sybyl* with an SPL (Sybyl Programming Language) program developed in our lab. Calculation of 3-D Wiener numbers consists of the sum entries in the upper triangular submatrix of the topographic Euclidean distance matrix for a molecule. The 3-D coordinates for the atoms were determined using *CONCORD 3.0.1*.[65] Two variants of the 3-D Wiener number were calculated. For ${}^{3D}W_H$, hydrogen atoms are included in the computations, and for ${}^{3D}W$, hydrogen atoms are excluded from the computations.

2.4. Data Reduction. Initially, all TIs were transformed by the natural logarithm of the index plus one. This was done since the scale of some indices may be several orders of magnitude greater than that of other indices. This scaling was also done for the geometric indices for consistency.

The set of 92 TIs was divided into two distinct sets: topostructural indices (TSI) and topochemical indices (TCI). TSIs are topological indices which encode information about the adjacency and distances of atoms (vertices) in molecular structures (graphs) irrespective of the chemical nature of the atoms involved in the bonding or factors such as hybridization states of atoms, number of core/valence electrons in individual atoms, etc. TCIs are parameters which quantify information regarding the topology (connectivity of atoms) as well as specific chemical properties of the atoms comprising a molecule. TCIs are derived from weighted molecular graphs where each vertex (atom) is properly weighted with relevant chemical/physical properties. Table 4 shows the breakdown of the topological indices into structural and chemical indices.

The sets of TSIs and TCIs were further divided into subsets, or clusters, based on the correlation matrix by using the SAS procedure VARCLUS.[66] The VARCLUS procedure divides the set of indices into disjoint clusters so that each cluster is essentially unidimensional.

From each cluster we selected the TI most correlated with the cluster as well as any TIs which were poorly correlated with the cluster ($R < 0.70$). These TIs were then used in the modeling of benzamidine-mediated inhibition of guinea pig complement. The variable clustering and selection of TIs was performed independently for both the TSI and TCI sets of indices.

2.5. Statistical Analysis. Regression modeling was accomplished using the SAS procedure REG.[66] During the initial stages of statistical analysis it became apparent that it

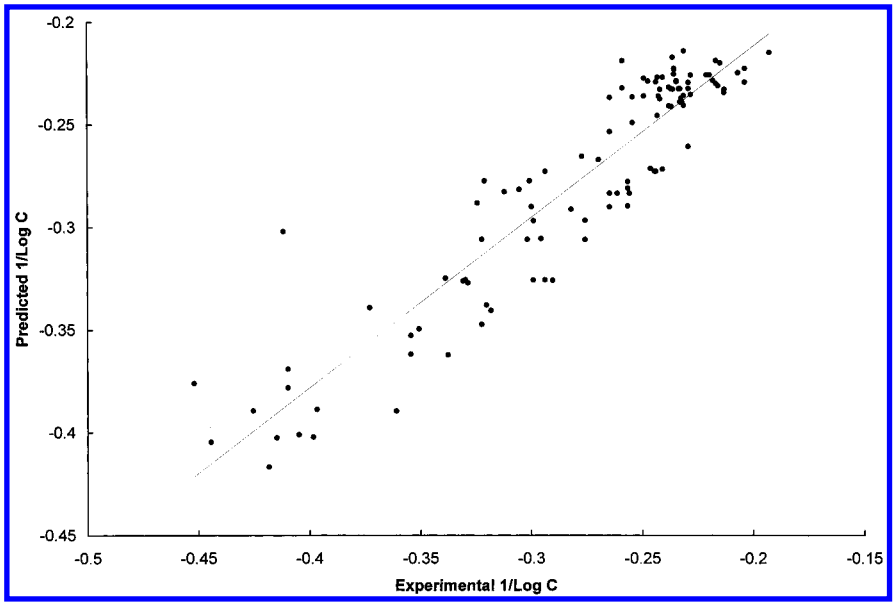

Figure 2. Scatterplot for observed 1/log *C* versus predicted 1/log *C* using eq 2 for the set of 107 benzamidines.

Table 4. Classification of Parameters Used in Developing Models for Complement Inhibition

topological	topochemical	geometric
I_D^W	I_{ORB}	V_w
$\overline{I_D^W}$	$IC_0 - IC_6$	^{3D}W
W	$SIC_0 - SIC_6$	$^{3D}W_H$
I^D	$ClC_0 - ClC_6$	
H^V	$^0\chi^b - {}^6\chi^b$	
H^D	$^0\chi^b{}_C - {}^6\chi^b{}_C$	
$\overline{IC}$	$^6\chi^b{}_{Ch}$	
O	$^4\chi^b{}_{PC} - {}^6\chi^b{}_{PC}$	
M_1	$^0\chi^v - {}^6\chi^v$	
M_2	$^0\chi^v{}_C - {}^6\chi^v{}_C$	
$^0\chi - {}^6\chi$	$^6\chi^b{}_{Ch}$	
$^3\chi_C - {}^6\chi_C$	$^4\chi^b{}_{PC} - {}^6\chi^b{}_{PC}$	
$^6\chi_{Ch}$	J^B	
$^4\chi_{PC} - {}^6\chi_{PC}$	J^X	
$P_0 - P_{10}$	J^Y	
J		

would be necessary to perform an alternative transformation of the data. Using Hansch and Yoshimoto's Log 1/*C* transformation resulted in residual plots that showed that the variance of the errors correlated with the predictions. To deal with this problem, we back transformed the data to the initial value *C* and then tried several other transformations, finally settling on 1/Log *C* which resulted in an uncorrelated residual plot. All subsets linear regression was then carried out on three distinct sets of indices: set I—three TSIs; set II—the TSI used in model I and four TCIs; and set III—the TSI retained in model I and the three geometrical indices. The regression analysis resulted in the final selection of TIs for the models.

3. RESULTS

Using only the topostructural class of indices, all-subsets regression resulted in a one parameter model to estimate I_{50}:

$$1/\log C = -1.1245 + 0.4989(I^D) \qquad (1)$$

$$n = 105,\ r = 0.940,\ r_c = 0.938,\ s = 0.0200,\ F = 785$$

This parameter was added to the set of topochemical parameters. Again, all-subsets regression was used to develop a model using this new set of independent variables. The best model for estimation of I_{50} once again used only I^D. This being the case, topochemical parameters were dropped from the modeling procedure.

Using all-subsets regression on the one parameter from eq 1 and the three geometrical parameters resulted in the selection of a different one parameter model:

$$1/\log C = -0.6428 + 0.0490(^{3D}W) \qquad (2)$$

$$n = 105,\ r = 0.943,\ r_c = 0.940,\ s = 0.0196,\ F = 824$$

Compounds 1 and 6 were removed from all models, as they were both strongly influential and were classified as outliers as defined by the studentized range. The predicted values from eq 2 for all 107 benzamidines, including the results predicted for the two outliers, are presented in Table 2.

A scatter plot of the experimental data for the 107 benzamidines versus the values predicted using eq 2 is presented in Figure 2. Predicted values for the two outliers have been included.

4. DISCUSSION

The objective of this paper was to study the relative effectiveness of topostructural, topochemical, and geometrical parameters in estimating the complement inhibitory potency of a set of benzamidines based solely on their chemical structures. Theoretical structural indices can be derived from distinct models of molecules. Also, various indices defined on the same representation of the molecule can quantify various aspects of molecular architecture. Recently, we have advocated the use of a "hierarchical QSAR approach" involving the TSI, TCI, geometrical, and quantum chemical indices in the successful development of predictive models.[67−71]

In comparing our study to the work of Hanch and Yoshimoto,[47] it must be pointed out that our models did little to improve on their QSAR analysis as can be seen from

examining our retransformed results with the results of their best equation.

	n	*r*	*s*
Basak et al.	105	0.943	0.264
Hansch and Yoshimoto	108	0.935	0.258

However, the LFER approach used by Hansch and Yoshimoto required experimental data for all compounds in the study and significant input from a human expert for the determination of the three "structural" indicator variables. One strength of our approach to this problem is the use of nonempirical theoretical descriptors which can be calculated solely from the chemical structure. With these purely theoretical descriptors we have modeled the inhibition of complement by benzamidines as successfully as Hansch and Yoshimoto using their LFER approach.

It is clear from this study of 107 benzamidines that the TSI indices are sufficient to explain most of the variance in bioactivity. The addition of TCI and geometrical parameters did not substantially increase the predictive power of the models. However, quantum chemical indices were not used for model development with this set of compounds.

TSIs encode information about generalized size and shape of a molecule. The success of TSI parameters in explaining most of the complement-inhibitory action of these benzamidines indicates that the general shape and size of these molecules largely determines their bioactivity. In some of our other studies we have found that the addition of quantum chemical indices can improve the correlation in cases of specific bioactivity. Further studies will focus on the contribution of quantum chemical indices in explaining the bioactivity of benzamidines.

ACKNOWLEDGMENT

This paper is contribution number 226 from the Center for Water and the Environment of the Natural Resources Research Institute. Research reported in this paper was supported, in part, by Grants F49620-94-1-0401 and F49620-96-1-0330 from the United States Air Force.

REFERENCES AND NOTES

(1) Basak, S. C.; Frane, C. M.; Rosen, M. E.; Magnuson, V. R. Molecular Topology and Acute Toxicity: A QSAR Study of Monoketones. *Med. Sci. Res.* **1987**, *15*, 887−888.

(2) Basak, S. C. Binding of Barbiturates to Cytochrome P_{450}: A QSAR Study Using Log P and Topological Indices. *Med. Sci. Res.* **1988**, *16*, 281−282.

(3) Charton, M. In *Steric Effects in Drug Design*; Charton, M., Motoc, M., Eds.; Springer-Verlag: Berlin, 1983; pp 107−118.

(4) Basak, S. C.; Niemi, G. J.; Veith, G. D. Predicting Properties of Molecules Using Graph Invariants. *J. Math. Chem.* **1991**, *7*, 243−272.

(5) Niemi, G. J.; Basak, S. C.; Veith, G. D.; Grunwald, G. Prediction of Octanol−Water Partition Coefficient (K_{OW}) Using Algorithmically-Derived Variables. *Environ. Toxicol. Chem.* **1992**, *11*, 893−900.

(6) Balaban, A. T.; Bertelsen, S.; Basak, S. C. New Centric Topological Indexes for Acyclic Molecules (Trees) and Substituents (Rooted Trees) and Coding of Rooted Trees. *Math. Chem.* **1994**, *30*, 55−72.

(7) Basak, S. C.; Bertelsen, S.; Grunwald, G. D. Use of Graph Theoretic Parameters in Risk Assessment of Chemicals. *Toxicol. Lett.* **1995**, *79*, 239−250.

(8) Balaban, A. T.; Basak, S. C.; Colburn, T.; Grunwald, G. Correlation Between Structure and Normal Boiling Points of Haloalkanes C_1−C_4 using Neural Networks. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1118−1121.

(9) Basak, S. C.; Grunwald, G. D. In *Proceeding of the XVI International Cancer Congress*, R. S. Rao, M. G. Deo, L. D. Sanghui, Eds.; Monduzzi: Bologna, Italy, 1995; pp 413−416.

(10) Basak, S. C. Use of Molecular Complexity Indices in Predictive Pharmacology and Toxicology: A QSAR Approach. *Med. Sci. Res.* **1987**, *15*, 605−609.

(11) Basak, S. C.; Niemi, G. J.; Veith, G. D. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; NOVA: New York, 1990; pp 235−277.

(12) Basak, S. C. In *Practical Applications of Quantitative Structure−Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devillers, J., Eds.; Kluwer Academic: Dordrecht/Boston/London, 1990; pp 83−103.

(13) Basak, S. C. In *Proceedings of the NATO Advanced Study Institute (ASI) on Pharmacokinetics*; Pub: Sicily, in press.

(14) Basak, S. C.; Grunwald, G. D.; Niemi, G. J. Use of Graph-Theoretic and Geometrical Molecular Descriptors in Structure−Activity Relationships. In *From Chemical Topology to Three-Dimensional Molecular Geometry*; Balaban, A. T., Ed.; Plenum Press: New York, 1997; pp 73−116.

(15) Basak, S. C.; Gute, B. D. Use of Graph Theoretic Parameters in Predicting Inhibition of Microsomal *p*-Hydroxylation of Anilines by Alcohol: A Molecular Similarity Approach. In *Proceedings of the International Congress on Hazardous Waste: Impact on Human and Ecological Health*; Johnson, B. L., Xintaras, C., Andrews, J. S., Jr., Eds.; Princeton Scientific Publishing Co., Inc.: Princeton, NJ, 1997; pp 492−504.

(16) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*; Research Studies Press: Letchworth, Hertfordshire, U.K, 1986.

(17) Basak, S. C.; Rosen, M. E.; Magnuson, V. R. Molecular Topology and Mutagenicity: A QSAR Study of Nitrosamines. *IRCS Med. Sci.* **1986**, *14*, 848−849.

(18) Basak, S. C.; Gieschen, D. P.; Magnuson, V. R.; Harriss, D. K. Structure−Activity Relationships and Pharmacokinetics: A Comparative Study of Hydrophobicity, van der Waals' Volume and Topological Parameters. *IRCS Med. Sci.* **1982**, *10*, 619−620.

(19) Basak, S. C.; Grunwald, G. D. Use of Graph Invariants, Volume and Total Surface Area in Predicting Boiling Point of Alkanes. *Mathl. Modelling Sci. Computing* **1993**, *2*, 735−740.

(20) Rouvray, D. H.; Pandey, R. B. The Fractal Nature, Graph Invariants and Physicochemical Properties of Normal Alkanes. *J. Chem. Phys.* **1986**, *85*, 2286−2290.

(21) Randić, M. Resolution of Ambiguities in Structure−Property Studies by Use of Orthogonal Descriptors. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311−320.

(22) Randić, M. Nonempirical Approaches to Structure−Activity Studies. *Int. J. Quantum Chem: Quantum Biol. Symp.* **1984**, *11*, 137−153.

(23) Trinajstić, N. *Chemical Graph Theory*, Klein, D. J., Randić, M., Eds.; CRC Press: Boca Raton, FL, 1992.

(24) Basak, S. C.; Bertelsen, S.; Grunwald, G. D. Application of Graph Theoretical Parameters in Quantifying Molecular Similarity and Structure−Activity Studies. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 270−276.

(25) Basak, S. C.; Grunwald, G. D. Molecular Similarity and Risk Assessment: Analogue Selection and Property Estimation Using Graph Invariants. *SAR QSAR Environ. Res.* **1994**, *2*, 289−307.

(26) Basak, S. C.; Gute, B. D.; Grunwald, G. D. Estimation of Normal Boiling Points of Haloalkanes Using Molecular Similarity. *Croat. Chim. Acta* **1996**, *69*, 1159−1173.

(27) Basak, S. C.; Magnuson, V. R.; Niemi, G. J.; Regal, R. R. Determining Structural Similarity of Chemicals using Graph-Theoretic Indices. *Discrete Appl. Math.* **1988**, *19*, 17−44.

(28) Basak, S. C.; Grunwald, G. D. Use of Topological Space and Property Space in Selecting Structural Analogues. *Mathl. Modelling Sci. Computing* **1998**, in press.

(29) Basak, S. C.; Grunwald, G. D. Estimation of Lipophilicity from Structural Similarity. *New J. Chem.* **1995**, *19*, 231−237.

(30) Basak, S. C.; Grunwald, G. D. Molecular Similarity and Estimation of Molecular Properties. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 366−372.

(31) Basak, S. C.; Grunwald, G. D. Tolerance Space and Molecular Similarity. *SAR QSAR Environ. Res.* **1995**, *3*, 265−277.

(32) Basak, S. C.; Grunwald, G. D. Predicting Mutagenicity of Chemicals Using Topological and Quantum Chemical Parameters: A Similarity Based Study. *Chemosphere* **1995**, *31*, 2529−2546.

(33) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom Pairs as Molecular Features in Structure−Activity Studies: Definition and Applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64−73.

(34) Johnson, M.; Basak, S. C.; Maggiora, G. A Characterization of Molecular Similarity Methods for Property Prediction. *Mathematical Comput. Modelling* **1988**, *II*, 630−635.

(35) Lajiness, M. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova: New York, 1990; pp 299−316.
(36) Wilkins, C. L.; Randić, M. A Graph Theoretic Approach to Structure−Property and Structure−Activity Correlations. *Theor. Chim. Acta (Berl.)* **1980**, *58*, 45−68.
(37) Basak, S. C.; Gute, B. D.; Grunwald, G. D. Development and Applications of Molecular Similarity Methods using Nonempirical Parameters. *Mathl. Modelling Sci. Computing*, in press.
(38) Fisanick, W.; Cross, K.; Ruzinko, III, A. Similarity Searching on CAS Registry Substances. 1. Global Molecular Property and Generic Atom Triangle Geometric Searching. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 664−674.
(39) Randić, M. Similarity Based on Extended Basis Descriptors. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 686−692.
(40) Mekenyan, O.; Peitchev, D.; Bonchev, D.; Trinajstić, N.; Bangov, I. Modelling the Interaction of Small Organic Molecules with Biomacromolecules. *Arzneim. Forsch.* **1986**, *36*, 176−183.
(41) Mihlic, Z.; Trinajstić, N. The Algebraic Modelling of Chemical Structures: On the Development of Three-Dimensional Molecular Descriptors. *J. Mol. Struct. (Theochem.)* **1991**, *232*, 65−78.
(42) Ray, S. K.; Basak, S. C.; Raychaudhury, C.; Roy, A. B.; Ghosh, J. J. The Utility of Information Content (IC), Structural Information Content (SIC), Hydrophobicity (log P) and van der Waals' Volume (Vw) in the Design of Barbiturates and Tumor-Inhibitory Triazenes: A Comparative Study. *Arzneim.-Forsch.* **1983**, *33*, 352−356.
(43) Ray, S. K.; Basak, S. C.; Raychaudhury, C.; Roy, A. B.; Ghosh, J. J. A Quantitative Structure−Activity Relationship (QSAR) Study of Barbiturates, Spasmolytics and Diphen-hydramines using van der Waals' Volume. *Acta Ciencia Indica* **1981**, *4*, 187−192.
(44) Koyama, M.; Ohtani, N.; Kai, F.; Moriguchi, I.; Inouye, S. Synthesis and Quantitative Structure−Activity Relationship Analysis of N-triiodoallyl and N-iodopropargylazole: New Antifungal Agents. *J. Med. Chem.* **1987**, *30*, 552−562.
(45) Lachmann, P. J. In *The Immune System*; Hobart, M. J., McConnell, I., Eds.; Blackwell Scientific Publications: Philadelphia, PA, 1976.
(46) Kuby, J. *Immunology*. W. H. Freeman & Co.: New York, 1992.
(47) Hansch, C.; Yoshimoto, M. Structure−Activity Relationships in Immunochemistry. 2. Inhibition of Complement by Benzamidines. *J. Med. Chem.* **1974**, *17*, 1160−1167.
(48) Baker, B. R.; Erickson, E. H. Irreversible Enzyme Inhibitors. CLII. Proteolytic Enzymes. X. Inhibition of Guinea Pig Complement by Substituted Benzamidines. *J. Med. Chem.* **1969**, *12*, 408−414.
(49) Baker, B. R.; Cory, M. Irreversible Enzyme Inhibitors. CLXIV. Proteolytic Enzymes. XIV. Inhibition of Guinea Pig Complement by *meta*-Substituted Benzamidines. *J. Med. Chem.* **1969**, *12*, 1049−1052.
(50) Baker, B. R.; Cory, M. Irreversible Enzyme Inhibitors. CLXV. Proteolytic Enzymes. XV. Inhibition of Guinea Pig Complement by Derivatives of m-Phenoxypropoxybenzamidine. *J. Med. Chem.* **1969**, *12*, 1053−1056.
(51) Baker, B. R.; Cory, M. Irreversible Enzyme Inhibitors. 180. Irreversible Inhibitors of the C′1a Component of Complement Derived from m-(Phenoxypropoxy) benzamidine and Phenoxyacetamide. *J. Med. Chem.* **1971**, *14*, 119−125.
(52) Baker, B. R.; Cory, M. Irreversible Enzyme Inhibitors. 186. Irreversible Inhibitors of the C′1a Component of Complement Derived from m-(Phenoxypropoxy) benzamidine by Bridging to a Terminal Sulfonyl Fluoride. *J. Med. Chem.* **1971**, *14*, 805−808.
(53) Basak, S. C.; Harriss, D. K.; Magnuson, V. R. POLLY 2.3; Copyright of the University of Minnesota, 1988.
(54) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.
(55) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.
(56) Raychaudhury, C.; Ray, S. K.; Ghosh, J. J.; Roy, A. B.; Basak, S. C. Discrimination of Isomeric Structures Using Information Theoretic Topological Indices. *J. Comput. Chem.* **1984**, *5*, 581−588.
(57) Bonchev, D.; Trinajstić, N. Information Theory, Distance Matrix and Molecular Branching. *J. Chem. Phys.* **1977**, *67*, 4517−4533.
(58) Basak, S. C.; Magnuson, V. R. Molecular Topology and Narcosis: A Quantitative Structure−Activity Relationship (QSAR) Study of Alcohols Using Complementary Information Content (CIC). *Arzneim. Forsch.* **1983**, *33*, 501−503.
(59) Basak, S. C.; Roy, A. B.; Ghosh, J. J. In *Proceedings of the IInd International Conference on Mathematical Modelling*; Avula, X. J. R., Bellman, R., Luke, Y. L., Rigler, A. K., Eds.; University of Missouri−Rolla: Rolla, MO, 1980; Vol. II, pp 851−856.
(60) Roy, A. B.; Basak, S. C.; Harriss, D. K.; Magnuson, V. R. In *Mathematical Modelling in Science and Technology*; Avula, X. J. R., Kalman, R. E., Lipais, A. I., Rodin, E. Y., Eds.; Pergamon Press: New York, 1984; pp 745−750.
(61) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399−404.
(62) Balaban, A. T. Topological Indices Based on Topological Distances in Molecular Graphs. *Pure Appl. Chem.* **1983**, *55*, 199−206.
(63) Balaban, A. T. Chemical Graphs. Part 48. Topological Index J for Heteroatom-Containing Molecules Taking into account Periodicities of Element Properties. *Math. Chem. (MATCH)* **1986**, *21*, 115−122.
(64) Tripos Associates, Inc. *SYBYL Version 6.1*; Tripos Associates, Inc.: St. Louis, MO, 1994.
(65) Tripos Associates, Inc. *CONCORD Version 3.0.1*; Tripos Associates, Inc.: St. Louis, MO, 1993.
(66) SAS Institute Inc. In *SAS/STAT User's Guide, Release 6.03 Edition*; SAS Institute Inc.: Cary, NC, 1988; Chapters 28 and 34, pp 773−875, 949−965.
(67) Basak, S. C.; Gute, B. D.; Grunwald, G. D. A Comparative Study of Topological and Geometrical Parameters in Estimating Normal Boiling Point and Octanol/Water Partition Coefficient, *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 1054−1060.
(68) Basak, S. C.; Gute, B. D.; Grunwald, G. D. Use of Topostructural, Topochemical and Geometric Parameters in the Prediction of Vapor Pressure: A Hierarchical QSAR Approach, *J. Chem. Inf. Comput. Sci*. **1997**, *37*, 651−655.
(69) Gute, B. D.; Basak, S. C. Predicting Acute Toxicity (LC_{50}) of Benzene Derivatives Using Theoretical Molecular Descriptors: A Hierarchical QSAR Approach. *SAR QSAR Environ. Res.* **1997**, *7*, 117−131.
(70) Basak, S. C.; Gute, B. D.; Grunwald, G. D. Relative Effectiveness of Topological, Geometrical, and Quantum Chemical Parameters in Estimating Mutagenicity of Chemicals, Quantitative Structure−Activity Relationships. In *Quantitative Structure−Activity Relationships in Environmental Sciences*; Chen, F., Schuurman, G., Eds.; SETAC Press: Pensacola, FL, 1997; Vol. 7, Chapter 17, p 245.
(71) Gute, B. D.; Grunwald, G. D.; Basak, S. C. Prediction of the Dermal Penetration of Polycyclic Aromatic Hydrocarbons (PAHs): A Hierarchical QSAR Approach, *SAR QSAR Environ. Res.* **1997**, in press.

CI9800716