

Use of topological indices in predicting aryl hydrocarbon receptor binding potency of dibenzofurans: A hierarchical QSAR approach

Subhash C Basak^{a*}, Denise Mills^a, Moiz M Mumtaz^b & Krishnan Balasubramanian^c

^a Natural Resources Research Institute, University of Minnesota Duluth, 5013 Miller Trunk Highway, Duluth, MN 55811, USA

^b Computational Toxicology Laboratory, Division of Toxicology, Agency for Toxic Substances and Disease Registry (ATSDR), Executive Park Building 4, 1600 Clifton Road, E-29, Atlanta, GA 30333, USA

^c Department of Applied Sciences, University of California Davis, Hertz Hall, L-794 Livermore, CA 94550, Chemistry & Material Science Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550 and Glenn T Seaborg Center, Lawrence Berkeley National Laboratory, Berkeley CA 94720

Received 1 February 2003

Topostructural (TS) and topochemical (TC) indices, geometrical descriptors, and *ab initio* (STO-3G) quantum chemical indices have been employed either alone or hierarchically in the development of quantitative structure-activity relationship (QSAR) models of the aryl hydrocarbon (Ah) receptor binding potency of a set of 34 dibenzofurans. Results show that, for the full set, the TS and TC indices explain most of the variance in the data. The addition of 3-D and quantum chemical indices makes only slight improvement in the predictive capability of QSAR models.

Introduction

An important problem in mathematical chemistry, drug discovery and computational toxicology is the prediction of activity/ toxicity/ property of molecules from their structure¹⁻⁵. In pharmaceutical drug design, one usually begins the process with the discovery of a lead compound, then synthesizes and tests thousands of derivatives of the lead structure in order to find a useful drug. The process is a costly one. Part of the cost can be saved if the potential therapeutic activity and toxicity of the derivatives can be tested by *in silico* methods instead of the costly *in vivo* or *in vitro* methods. In environmental toxicology and human health hazard assessment of chemicals, there are many thousands of chemicals that need to be tested for their effects on human and ecological health. In the USA, the Toxic Substances Control Act (TSCA) Inventory currently has more than 82,000 chemicals, most of which have very little experimental data for the estimation of their potential hazard⁶. The American Chemistry Council has recently embarked on a plan to test approximately 3,000 high production volume (HPV) chemicals from the TSCA Inventory at a cost of nearly 700 million dollars. Combinatorial chemistry is producing many large real as well virtual libraries, which need to be evaluated for activity/toxicity on the fly. Most of these chemicals have virtually no experimental data at all.

A perusal of the above account clearly indicates that both in drug design and in hazard assessment of environmental pollutants there is a compelling need for data on various properties that are used in assessing potential therapeutic/toxic effects of molecules. Very limited resources and testing facilities are available for testing the large pool of candidate chemicals that we have to deal with for drug design as well as protection of ecological and human health. A viable solution to this quagmire has been the estimation of necessary properties of molecules directly from their structure without the input of any other experimental data via quantitative structure-activity relationship (QSAR) models. Numerous QSAR studies of recent years have used topological indices in the development of QSARs pertaining to medicinal chemistry and environmental toxicology^{1,3,7-11}.

A graph $G = (V,E)$ represents a molecule when the vertex set V represents the set of atoms and the edge set E symbolizes the set of bonds, usually covalent bonds. Such a graph represents the topology of the chemical species in the sense that a molecular graph preserves the connectedness of atoms in the molecule, at the same time being independent of the metric aspects of the molecular structure such as bond angle, bond length, etc. Once a chemical is represented by a molecular graph, its structure can be characterized by

graph invariants. A graph invariant is a graph theoretic property that is the same or has the same value for isomorphic graphs. A graph invariant can be a polynomial (e.g., the characteristic polynomial), a set of numbers (e.g. the spectra of a graph) or a single number. A single number that characterizes the topology of a molecular graph has been termed a topological index by Hosoya¹². Many topological indices (TIs) have already been described in the literature^{3,13-15}. In the past two decades, TIs have been used in isomer discrimination¹⁶, characterization of closely related structures¹⁷, ordering and partial ordering of chemicals^{18,19}, and QSAR studies to predict property/ activity/ toxicity of chemicals.

We have recently formulated a hierarchical QSAR (HiQSAR) approach in which increasingly complex and computationally demanding parameters are used in a graduated manner. The first level is comprised of the topostructural (TS) indices, which encode information about the adjacency and distance between vertices in a molecular graph. The topochemical (TC) indices, which constitute the second level, consist of topological indices that quantify information not only regarding the connectivity, but also about the chemical nature of atoms or bonds. At the third and fourth levels we have the 3-D and quantum chemical (QC) indices that are used only when the TS and TC indices do not give predictive models of acceptable quality. The HiQSAR approach has been successfully applied in the prediction of many properties including complement inhibitory activity of benzamides²⁰, mutagenicity of 95 aromatic amines^{21, 22}, mutagenicity/ non-mutagenicity of a set of 508 diverse chemicals²³, vapor pressure of a diverse group of 476 molecules^{24, 25}, boiling point of a structurally very diverse set of 1,015 chemicals²⁶, and the cellular toxicity of a set of 55 halocarbons²⁷.

Dibenzofurans are widespread environmental contaminants that are produced mainly as undesirable by-products in natural and industrial processes. The toxic effects of these compounds are thought to be mediated through binding to the aryl hydrocarbon (Ah) receptor. In this paper we have used our HiQSAR approach in the development of QSAR models to predict aryl hydrocarbon (Ah) receptor binding potency utilizing a set of 34 dibenzofurans.

Materials and Methods

Database

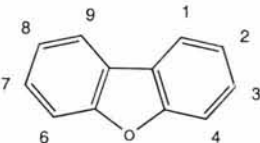
Aryl hydrocarbon (Ah) receptor binding affinity data for a set of 34 chlorinated dibenzofurans were obtained from the literature²⁸. The activity is reported

as pEC₅₀. Compound names and observed Ah receptor binding affinities are provided in Table 1.

Molecular descriptors

Software programs, including POLLY v2.3²⁹, Triplet³⁰, Molconn-Z v3.50³¹, and Gaussian 98W v5.1³², were used in the calculation of molecular descriptors, all of which are derived from chemical structure. The descriptors are partitioned into four classes: Topostructural (TS), topochemical (TC), geometrical (3D), and *ab initio* quantum chemical (QC). TS descriptors encode information strictly on

Table 1—Structure and Ah receptor binding affinity data for dibenzofuran and chlorinated derivatives.

No.	Chemical	Observed pEC ₅₀ ^a	Predicted pEC ₅₀ ^b
			
1	2-Cl	3.553	3.16905
2	3-Cl	4.377	4.19880
3	4-Cl	3.000	3.69217
4	2,3-diCl	5.326	4.96434
5	2,6-diCl	3.609	4.27872
6	2,8-diCl	3.590	4.25137
7	1,2,7-trCl	6.347	5.64627
8	1,3,6-trCl	5.357	4.70494
9	1,3,8-trCl	4.071	5.33036
10	2,3,4-trCl	4.721	----
11	2,3,8-trCl	6.000	6.39401
12	1,2,3,6-teCl	6.456	6.47979
13	1,2,3,7-teCl	6.959	7.06574
14	1,2,4,8-teCl	5.000	4.71451
15	2,3,4,6-teCl	6.456	7.32118
16	2,3,4,7-teCl	7.602	7.49601
17	2,3,4,8-teCl	6.699	6.97567
18	2,3,6,8-teCl	6.658	6.00843
19	2,3,7,8-teCl	7.387	7.13937
20	1,2,3,4,8-peCl	6.921	6.29270
21	1,2,3,7,8-peCl	7.128	7.21285
22	1,2,3,7,9-peCl	6.398	5.72435
23	1,2,4,6,7-peCl	7.169	6.13450
24	1,2,4,6,8-peCl	5.509	-----
25	1,2,4,7,8-peCl	5.886	6.60650
26	1,2,4,7,9-peCl	4.699	4.93707
27	1,3,4,7,8-peCl	6.699	6.51315
28	2,3,4,7,8-peCl	7.824	7.47861
29	2,3,4,7,9-peCl	6.699	6.50924
30	1,2,3,4,7,8-heCl	6.638	6.80214
31	1,2,3,6,7,8-heCl	6.569	7.12358
32	1,2,4,6,7,8-heCl	5.081	5.67190
33	2,3,4,6,7,8-heCl	7.328	7.01939
34	Dibenzofuran	3.000	2.76503

^aObserved values obtained by So and Karpulus²⁸

^bCross-validated results obtained using TS+TC ridge regression model, based on 32 compounds

the adjacency and topological distances between the atoms within a molecule, while TC descriptors also take into account chemical information such as bond and atom types. TS and TC are collectively referred to as topological descriptors as they are based on molecular topology. The 3D descriptors encode information on the geometrical aspects of molecular structure, and the QC descriptors encode the electronic aspects of molecular structure. These descriptor classes can be ordered as follows with respect to their complexity and demand for computational resources: TS < TC << 3D <<< QC. As such, it was of interest to determine whether the simple parameters are adequate to produce good quality predictive models or if the more time intensive parameters are required. The topological descriptors were obtained from POLLY, Triplet, and Molconn-Z. Examples include the connectivity indices³³⁻³⁵, path length descriptors³⁵, descriptors of polarity and hydrogen bonding potential, electropological state indices^{36,37}, and information theoretic and neighborhood complexity indices³⁸⁻⁴¹. In addition to a large number of topological indices, Molconn-Z also provides us with the 3D descriptors used in this study, specifically, a set of kappa shape indices. Gaussian 98W was used to optimize the geometry of each molecule in the data set with Hartree-Fock calculations using the 3-21G basis set, and subsequently calculate the *ab initio* quantum chemical descriptors at the STO-3G level. A complete list of the molecular descriptors calculated for use in this study is provided in Table 2.

In the hierarchical quantitative structure-activity relationship (HiQSAR) approach, we create multiple models based on the inclusion of progressively more complex descriptors in an effort to gain insight into the relative contributions of each descriptor class.

Statistical analysis

All descriptors, other than the *ab initio* quantum chemical, were transformed by the natural logarithm as their scales differed by several orders of magnitude. The CORR procedure of the SAS statistical package⁴² was used to identify perfectly correlated descriptors. In each case, only one descriptor of each such pair was retained for use in the subsequent analysis. Any descriptor with a constant value for all chemicals in the data set was eliminated from the study.

For comparative purposes, three linear regression modeling methods were used; namely, ridge regression (RR)^{43,44}, principal components regression

(PCR)⁴⁵, and partial least squares (PLS)⁴⁶. Each of these methods is useful when the number of independent variables exceeds the number of observations and/or when the independent variables are highly intercorrelated. The statistical measures reported include the cross-validated R² and the PRESS statistic, both reliable measures of model predictability.

To derive the cross-validated R², each compound in turn is omitted from the data set, and the coefficients of the regression model (RR, PLS or PCR) are computed using the remaining n-1 cases. These coefficients are then used to predict the held-out case. A cross-validation R² can be defined by

$$R_{cv}^2 = 1 - \frac{PRESS}{SSTotal}$$

where *PRESS* is the predictive residual sum of squares and *SSTotal* is the total sum of squares.

It is important to distinguish between a fitted and cross-validated R². The addition of any descriptor in a regression study, including those that are irrelevant, will tend to *increase* the value of the fitted R², resulting in a statistical measure that can be quite misleading. However, the addition of irrelevant descriptors tends to *decrease* the value of the cross-validated R², providing a much more realistic measure of model predictability.

Results and Discussion

Statistical results are provided in Table 3. Generally, RR produced better models than either PCR or PLS. Examining the hierarchically developed models in the upper half of Table 3, we find that the addition of the TC to the TS descriptors does not result in an improved model. This is contrary to our typical findings with hierarchical QSAR where we have found significant improvement in model quality when the TC descriptors are added to the TS^{21,22,24,25,47-50}. The addition of the 3D descriptors does not result in improvement in model quality, but the addition of the *ab initio* QC descriptors does result in minimal improvement. Examining the one-class models in the lower half of the table, we find that the TS and TC descriptors outperform the 3D and QC descriptors.

Continued statistical analysis revealed 2,3,4-trichlorodibenzofuran and 1,2,4,6,8-pentachlorodibenzofuran to be highly influential and, as such, they were removed from the data set. A summary of the

Table 2—Symbols, definitions and classification of calculated molecular descriptors

<i>Topostructural (TS)</i>	
I_D^W	Information index for the magnitudes of distances between all possible pairs of vertices of a graph
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
IC	Information content of the distance matrix partitioned by frequency of occurrences of distance h
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
${}^h\chi$	Path connectivity index of order h = 0-10
${}^h\chi_C$	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 = 3-10
P_h	Number of paths of length h = 0-10
J	Balaban's J index based on topological distance
nrings	Number of rings in a graph
ncirc	Number of circuits in a graph
DN^2S_y	Triplet index from distance matrix, square of graph order (# of non-H atoms), and distance sum; operation y = 1-5
DN^2I_y	Triplet index from distance matrix, square of graph order, and number 1; operation y = 1-5
$AS1_y$	Triplet index from adjacency matrix, distance sum, and number 1; operation y = 1-5
$DS1_y$	Triplet index from distance matrix, distance sum, and number 1; operation y = 1-5
ASN_y	Triplet index from adjacency matrix, distance sum, and graph order; operation y = 1-5
DSN_y	Triplet index from distance matrix, distance sum, and graph order; operation y = 1-5
DN^2N_y	Triplet index from distance matrix, square of graph order, and graph order; operation y = 1-5
ANS_y	Triplet index from adjacency matrix, graph order, and distance sum; operation y = 1-5
ANI_y	Triplet index from adjacency matrix, graph order, and number 1; operation y = 1-5
ANN_y	Triplet index from adjacency matrix, graph order, and graph order again; operation y = 1-5
ASV_y	Triplet index from adjacency matrix, distance sum, and vertex degree; operation y = 1-5
DSV_y	Triplet index from distance matrix, distance sum, and vertex degree; operation y = 1-5
ANV_y	Triplet index from adjacency matrix, graph order, and vertex degree; operation y = 1-5
<i>Topochemical (TC)</i>	
O	Order of neighborhood when IC_r reaches its maximum value for the hydrogen-filled graph
O_{orb}	Order of neighborhood when IC_r reaches its maximum value for the hydrogen-suppressed graph
I_{orb}	Information content or complexity of the hydrogen-suppressed graph at its maximum neighborhood of vertices
IC_r	Mean information content or complexity of a graph based on the r^{th} ($r = 0-6$) order neighborhood of vertices in a hydrogen-filled graph
SIC_r	Structural information content for r^{th} ($r = 0-6$) order neighborhood of vertices in a hydrogen-filled graph
CIC_r	Complementary information content for r^{th} ($r = 0-6$) order neighborhood of vertices in a hydrogen-filled graph
${}^h\chi^b$	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 = 3-6
${}^h\chi_{PC}^b$	Bond path-cluster connectivity index of order h = 4-6
${}^h\chi^v$	Valence path connectivity index of order h = 0-10
${}^h\chi_C^v$	Valence cluster connectivity index of order h = 3-6
${}^h\chi_{Ch}^v$	Valence chain connectivity index of order h = 3-10
${}^h\chi_{PC}^v$	Valence path-cluster connectivity index of order h = 4-6
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
AZV_y	Triplet index from adjacency matrix, atomic number, and vertex degree; operation y = 1-5
AZS_y	Triplet index from adjacency matrix, atomic number, and distance sum; operation y = 1-5
ASZ_y	Triplet index from adjacency matrix, distance sum, and atomic number; operation y = 1-5
AZN_y	Triplet index from adjacency matrix, atomic number, and graph order; operation y = 1-5
ANZ_y	Triplet index from adjacency matrix, graph order, and atomic number; operation y = 1-5
DSZ_y	Triplet index from distance matrix, distance sum, and atomic number; operation y = 1-5

Contd—

Table 2—Symbols, definitions and classification of calculated molecular descriptors—*Contd*

DN ² Z _y	Triplet index from distance matrix, square of graph order, and atomic number; operation y = 1-5
nvx	Number of non-hydrogen atoms in a molecule
nelem	Number of elements in a molecule
fw	Molecular weight
si	Shannon information index
totop	Total Topological Index t
sumI	Sum of the intrinsic state values I
sumdelI	Sum of delta-I values
tets2	Total topological state index based on electrotopological state indices
phia	Flexibility index (kp1* kp2/nvx)
IdCbar	Bonchev-Trinajstic information index
IdC	Bonchev-Trinajstic information index
Wp	Wienerp
Pf	Plattf
Wt	Total Wiener number
knotp	Difference of chi-cluster-3 and path/cluster-4
knotpv	Valence difference of chi-cluster-3 and path/cluster-4
nclass	Number of classes of topologically (symmetry) equivalent graph vertices
numHBd	Number of hydrogen bond donors
numwHBd	Number of weak hydrogen bond donors
numHBa	Number of hydrogen bond acceptors
SHCsats	E-State of C sp ³ bonded to other saturated C atoms
SHCsatu	E-State of C sp ³ bonded to unsaturated C atoms
SHvin	E-State of C atoms in the vinyl group, =CH-
SHTvin	E-State of C atoms in the terminal vinyl group, =CH ₂
SHavin	E-State of C atoms in the vinyl group, =CH-, bonded to an aromatic C
SHarom	E-State of C sp ² which are part of an aromatic system
SHHBd	Hydrogen bond donor index, sum of Hydrogen E-State values for -OH, =NH, -NH ₂ , -NH-, -SH, and #CH
SHwHBd	Weak hydrogen bond donor index, sum of C-H Hydrogen E-State values for hydrogen atoms on a C to which a F and/or Cl are also bonded
SHHBa	Hydrogen bond acceptor index, sum of the E-State values for -OH, =NH, -NH ₂ , -NH-, >N-, -O-, -S-, along with -F and -Cl
Qv	General Polarity descriptor
NHBint _y	Count of potential internal hydrogen bonders (y = 2-10)
SHBint _y	E-State descriptors of potential internal hydrogen bond strength (y =2-10)
	Electrotopological State index values for atoms types: SHsOH, SHdNH, SHsSH, SHsNH ₂ , SHssNH, SHtCH, SHother, SHCHnX, Hmax Gmax, Hmin, Gmin, Hmaxpos, Hminneg, SsLi, SssBe, SssssBem, SssBH, SssB, SssssBm, SsCH ₃ , SdCH ₂ , SssCH ₂ , StCH, SdsCH, SaaCH, SssssCH, SddC, StsC, SdssC, SaasC, SaaaC, SssssC, SsNH ₃ p, SsNH ₂ , SssNH ₂ p, SdNH, SssNH, SaaNH, StN, SssssNHp, SdsN, SaaN, SssssN, SddsN, SaasN, SssssNp, SsOH, SdO, SsO, SaaO, SsF, SsSiH ₃ , SssSiH ₂ , SssssSiH, SssssSi, SsPH ₂ , SssPH, SssP, SdssP, SssssP, SsSH, SdS, SssS, SaaS, SdssS, SddssS, SssssssS, SsCl, SsGeH ₃ , SssGeH ₂ , SssssGeH, SssssGe, SsAsH ₂ , SssAsH, SssssAs, SdssAs, SssssAs, SsSeH, SdSe, SssSe, SaaSe, SdssSe, SddssSe, SsBr, SsSnH ₃ , SssSnH ₂ , SssssSnH, SssssSn, SsI, SsPbH ₃ , SssPbH ₂ , SssssPbH, SssssPb
	<i>Geometrical / Shape (3D)</i>
kp0	Kappa zero
kp1-kp3	Kappa simple indices
ka1-ka3	Kappa alpha indices
	<i>Ab Initio Quantum Chemical (QC)</i>
E _{HOMO}	Energy of the highest occupied molecular orbital
E _{HOMO-1}	Energy of the second highest occupied molecular orbital
E _{LUMO}	Energy of the lowest unoccupied molecular orbital
E _{LUMO+1}	Energy of the second lowest unoccupied molecular orbital
ΔE	HOMO-LUMO energy gap
μ	Dipole moment

results based on the remaining 32 compounds is provided in Table 4. Again we find better models produced by RR as opposed to either PCR or PLS, and little or no improvement in model quality when

the more complex and computationally demanding 3D and QC descriptors are added to the simpler TS and TC descriptors. However, there is model improvement when the TC descriptors are added to

Table 3—Summary statistics for predictive models, N = 34.

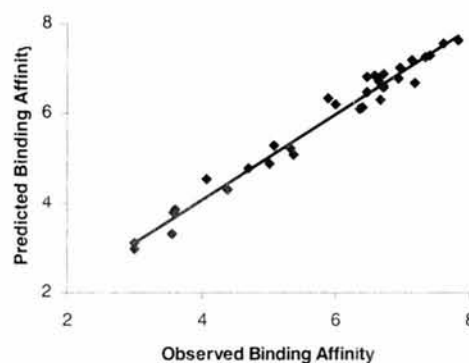
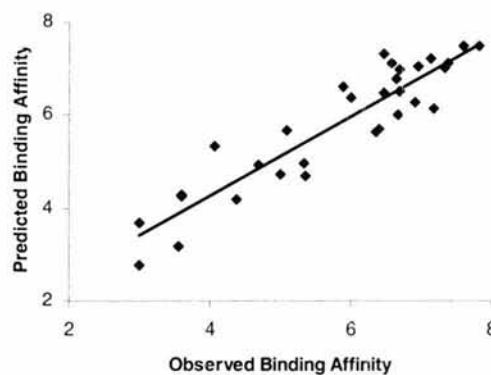
Model Type	RR		PCR		PLS	
	R ² _{c.v.}	PRESS	R ² _{c.v.}	PRESS	R ² _{c.v.}	PRESS
TS	0.691	19.7	0.662	21.6	0.631	23.6
TS+TC	0.675	20.8	0.623	24.1	0.468	34.0
TS+TC+3D	0.676	20.7	0.622	24.1	0.468	34.0
TS+TC+3D+STO-3G	0.725	17.6	0.575	27.2	0.564	27.9
TS	0.691	19.7	0.662	21.6	0.631	23.6
TC	0.696	19.4	0.622	24.1	0.531	30.0
3D	0.496	32.2	0.484	33.0	0.464	34.3
STO-3G	0.524	30.4	0.471	33.8	0.481	33.1

Table 4—Summary statistics for predictive models, N = 32.

Model Type	RR		PCR		PLS	
	R ² _{c.v.}	PRESS	R ² _{c.v.}	PRESS	R ² _{c.v.}	PRESS
TS	0.731	16.9	0.690	19.4	0.701	18.7
TS+TC	0.852	9.27	0.683	19.9	0.836	10.3
TS+TC+3D	0.852	9.27	0.683	19.9	0.837	10.2
TS+TC+3D+STO-3G	0.862	8.62	0.595	25.4	0.862	8.67
TS	0.731	16.9	0.690	19.4	0.701	18.7
TC	0.820	11.3	0.694	19.1	0.749	15.7
3D	0.508	30.8	0.523	29.9	0.419	36.4
STO-3G	0.544	28.6	0.458	33.9	0.501	31.3

the TS. As we found with the full set of 34 compounds, the best single-class model is that produced with the TC descriptors. Table 1 includes cross-validated predicted *Ah* receptor binding affinity values derived from the TS+TC ridge regression model based on 32 compounds, and Figs 1 and 2 represent the scatter plots of the fitted and cross-validated results, respectively.

The aim of this paper was to investigate the utility of the various classes of calculated indices in the formulation of predictive QSAR models. The results indicate that the topological indices (TS and TC parameters) explain most of the variance in the data. Our previous studies with various types of properties support such a conclusion. Topological indices, the TS encoding information regarding molecular size, shape, and branching characteristics, have been found to be strongly correlated with toxicologically relevant properties such as hydrophobicity and other molecular properties³⁵. This could be the reason behind the strong correlation between the TIs and *Ah* receptor binding affinity. It is expected that TIs, which require minimal computational resources, will find wide application in the estimation of property/ activity/ toxicity of molecules for which test data are scanty or totally unavailable.

Fig. 1—Observed vs predicted *Ah* receptor binding affinity (N=32, fitted TS+TC model)Fig. 2—Observed vs predicted *Ah* receptor binding affinity (N=32, cross-validated TS+TC model)

Acknowledgements

This is Contribution Number 331 from the Center for Water and the Environment of the Natural Resources Research Institute. Research reported in this paper was supported by Grant F49620-01-0098 from the United States Air Force and Grant/Cooperative Agreement Number 572112 from ATSDR. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of ATSDR. The work at UC Davis was supported by the National Science Foundation under Grant No. CHE-0236434. The work at LLNL was performed under the auspices of US Department of Energy by university of California under Contract No. W-7405-Eng-48.

References

- 1 Eldred D V & Jurs P C, *SAR QSAR environ Res*, 10 (1999) 75.
- 2 Klopman G & Tu M, *J med Chem*, 42 (1999) 992.
- 3 Devillers J & Balaban A T (eds), *Topological indices and related descriptors in QSAR and QSPR* (Gordon & Breach, Amsterdam), 1999.
- 4 Basak S C, Mills D, Gute B D, Grunwald G D & Balaban A T, in: Rouvray D H & King R B (eds), *Topology in chemistry: Discrete mathematics of molecules* (Horwood, Chichester), 2002, pp 113-184.
- 5 Balaban A T (ed), *From chemical topology to three-dimensional geometry* (Plenum Press, New York), 1997.
- 6 Auer C M, Zeeman M, Nabholz J V & Clements R G, *SAR QSAR environ Res*, 2 (1994) 29.
- 7 Eldred D V, Weikel C L, Jurs P C & Kaiser K L E, *Chem Res Toxicol*, 12 (1999) 670.
- 8 Serra R, Jurs P C & Kaiser K L E, *Chem Res Toxicol*, 14 (2001) 1535.
- 9 Blake B W, Enslein K, Gombar V K & Borgstedt H H, *Mutation Res*, 241 (1990) 261.
- 10 (Basak S C in: Karcher W & Devillers J (eds), *Practical applications of quantitative structure-activity relationships (QSAR) in environmental chemistry and toxicology* (Kluwer, Dordrecht), 1990, pp 83-103.
- 11 Basak S C, *Med Sci Res*, 15 (1987) 605.
- 12 Hosoya H, *Bull chem Soc Japan*, 44 (1971) 2332.
- 13 Gutman I, Kennedy J W & Quintas L V, *Chem Phys Lett.*, 173 (1990) 403.
- 14 Gutman I & Ivic A, *Discrete Math*, 150 (1996) 131.
- 15 Trinajstić N, *Chemical graph theory, Vols. I & II* (CRC Press, Boca Raton), 1983.
- 16 Raychaudhury C, Ray S K, Ghosh J J, Roy A B & Basak S C, *J comput Chem*, 5 (1984) 581.
- 17 Balasubramanian K & Basak S C, *J chem Inf Comput Sci*, 38 (1998) 367.
- 18 Randić M, Vračko M, Nović M & Basak S C, *MATCH Commun math Comput Chem*, 42 (2000) 181.
- 19 Basak S C, Magnuson V R, Niemi G J & Regal R R, *Discrete appl Math*, 19 (1988) 17.
- 20 Basak S C, Gute B D & Ghatak S, *J chem Inf Comput Sci*, 39 (1999) 255.
- 21 Basak S C, Gute B D & Grunwald G D, *SAR QSAR environ Res*, 10 (1999) 117.
- 22 Basak S C, Mills D R, Balaban A T & Gute B D, *J chem Inf Comput Sci*, 41 (2001) 671.
- 23 Basak S C, Mills D, Gute B D & Hawkins D M in: Benigni R (ed), *Quantitative structure-activity relationship (QSAR) models of mutagens and carcinogens* (CRC Press, Boca Raton), 2003, pp 207-234.
- 24 Basak S C, Gute B D & Grunwald G D, *J chem Inf Comput Sci*, 37 (1997) 651.
- 25 Basak S C & Mills D, *J chem Inf Comput Sci*, 41 (2001) 692.
- 26 Basak S C & Mills D, *MATCH Commun math Comput Chem*, 44 (2001) 15.
- 27 Basak S C, Balasubramanian K, Gute B D, Mills D, Gorczynska A & Roszak S, *J chem Inf Comput Sci*, (in press)
- 28 So S S & Karplus M, *J med Chem*, 40 (1997) 4360.
- 29 Basak S C, Harriss D K & Magnuson V R, POLLY, Version 2.3, Copyright Univ. Minnesota (1988).
- 30 Filip P A, Balaban T S & Balaban A T, *J math Chem*, 1 (1987) 61.
- 31 Molconn-Z, v 3.50, Hall Associates Consulting: Quincy, MA, 2000.
- 32 Gaussian 98W, v5.1 rev.A.6. ed.: Gaussian, Inc.:Carnegie, PA., 1998.
- 33 Randić M, *J Am chem Soc*, 97 (1975) 6609.
- 34 Kier L B, Murray W J, Randić M & Hall L H, *J pharm Sci*, 65 (1976) 1226.
- 35 Kier L B & Hall L H, *Molecular connectivity in structure-activity analysis* (Research Studies Press, Letchworth), 1986.
- 36 Kier L B & Hall L H, *Pharm Res*, 8 (1990) 801.
- 37 Kier L B, Hall L H & Frazer J W, *J math Chem*, 7 (1991) 229.
- 38 Bonchev D, *Information theoretic indices for characterization of chemical structures* (Research Studies Press, Letchworth), 1983.
- 39 Basak S C, Roy A B & Ghosh J J, Rolla, Missouri 1979; University of Missouri-Rolla; 851-856.
- 40 Roy A B, Basak S C, Harriss D K & Magnuson V R in: Avula X J R, Kalman R E, Liapis A I & Rodin E Y (eds), *Mathematical modelling in science and technology* (Pergamon, New York), 1984, pp 745-750.
- 41 Magnuson V R, Harriss D K & Basak S C in: King R B (ed), *Chemical applications of topology and graph theory* (Elsevier, Amsterdam), 1983, pp 178-191.
- 42 SAS Institute Inc.: Cary, NC., 1988.
- 43 Hoerl A E & Kennard R W, *Technometrics*, 8 (1970) 27.
- 44 Hawkins D M & Yin X, *Comput Statist Data Anal*, 40 (2002) 253.
- 45 Massy W F, *J Am statistic Assoc*, 60 (1965) 234.
- 46 Hoskuldsson A, *J Chemometrics*, 2 (1988) 211.
- 47 Gute B D & Basak S C, *SAR QSAR environ Res*, 7 (1997) 117.
- 48 Gute B D, Grunwald G D & Basak S C, *SAR QSAR environ Res*, 10 (1999) 1.
- 49 Basak S C, Gute B D & Grunwald G D in: Devillers J & Balaban A T (eds), *Topological indices and related descriptors in QSAR and QSPR* (Gordon & Breach, Amsterdam), 1999, pp 675-696.
- 50 Basak S C & Mills D, *SAR QSAR environ Res*, 12 (2001) 481.