Chemical Similarity Assessment through Multilevel Neighborhoods of Atoms: Definition and Comparison with the Other Descriptors

Dmitrii Filimonov, Vladimir Poroikov,* Yulia Borodina, and Tatyana Gloriozova

Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Pogodinskaya Street, 10, 119832 Moscow, Russia

Received September 10, 1998

A new method for assessment of molecular similarity based on original description of chemical structure is discussed. The accuracy of similarity assessment obtained with this method is compared with that of the results of four other approaches. The same evaluation set is used to predict: (a) boiling point of 139 hydrocarbons and (b) mutagenicity of 15 nitrosamines. The results show that the proposed method provides reasonable appraisal for both properties, but prediction of mutagenicity is more accurate in this method as compared to the alternatives.

INTRODUCTION

Assessment of chemical similarity is widely used in computer-aided study of new pharmaceuticals.[1,2] There are many ways to define the similarity of molecules using various structure descriptions and different mathematical methods for computing the similarity estimates.[3] Two groups of descriptors are traditionally used for similarity estimation: fragment substructures[4−6] and topological theoretic indices.[7]

Two-dimensional (2D) or three-dimensional (3D) fragment substructures are used for similarity assessment. Though 3D similarity methods are widely discussed in the literature,[3] 2D methods are still more suitable for analysis of large databases.[5,8]

Here we propose a new method for the analysis of similarity based on 2D description of molecules—multilevel neighborhoods of atoms (MNA). We have designed MNA as universal descriptors. About 400 types of biological activity are predicted by computer system PASS with more than 80% of mean accuracy estimated in leave one out cross-validation.[9−14] Such accuracy is achived by SAR analysis because each biological activity is qualitatively presented in PASS. In this paper we study the applicability of MNA descriptors for the assessment of quantitative characteristics such as boiling point, mutagenicity, etc. (QSAR/QSRP). Two sets of data (boiling point of hydrocarbons and mutagenicity of nitrosoamines) used before for the accuracy estimation in several other approaches[15,16] are feasible for comparative evaluation of MNA descriptors.

METHODS

Definition of Descriptors. The structure of a molecule is determined by the nature of its constituent atoms and the way that they are joined to one another. But its representation in a computer is conventional. Some ambiguty may occur when equivalent structures are drawn in various ways even with standard chemical editor. For example, the same substance (1,2-di-chlorobenzene) made by ISIS/Draw 2.0 (MDL Information Systems, Inc.) with different alternating single and double bonds produces different chains: Cl−C−C−Cl and Cl−C═C−Cl (see the bonds marked as bold in connection table given in Figure 1).

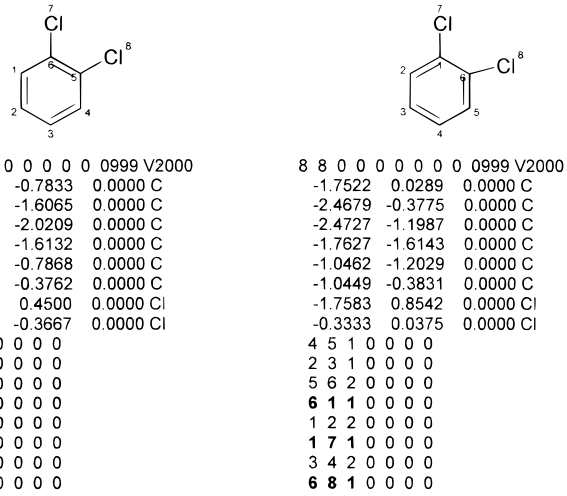


Figure 1. Two various displayed images of 1,2-dichlorobenzene and the difference in respective connection tables generated by ISIS/Draw 2.0 (the chain Cl−C−C−Cl is marked in bold).

For this reason, in contrast to many other widely used methods,[3] MNA descriptors are based on structure representation which does not specify the bond types and includes the hydrogens according to valencies and partial charges of atoms.

MNA descriptors are generated iteratively in the following way. The zero-level descriptor is presented by the type of atom according to Table 1. A special mark "-" is added to the descriptor of zero level if the atom is not included in the cycle. The descriptor of each successive level is a concatenation of the zero-level descriptor of the atom and, enclosed in parentheses, a lexicographically ordered list of descriptors of the previous level of its nearest neighbors. It can be shown that MNA descriptors are the analogues of the members in the formal series of Green's function for a molecule in quantum chemistry.[17]

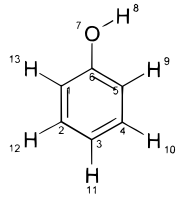
* To whom correspondence should be addressed. E-mail: vvp@ibmh.msk.su.

10.1021/ci980335o CCC: $18.00
Published on Web 06/10/1999

Table 1. Specification of Different Types of Atoms for Calculation of MNA Descriptors

type of atom	elements
H	H
C	C
N	N
O	O
F	F
Si	Si
P	P
S	S
Cl	Cl
Ca	Ca
As	As
Se	Se
Br	Br
Li*	Li, Na
B*	B, Re
Mg*	Mg, Mn
Sn*	Sn, Pb
Te*	Te, Po
I*	I, At
Os*	Os, Ir
Sc*	Sc, Ti, Zr
Fe*	Fe, Hf, Ta
Co*	Co, Sb, W
Sr*	Sr, Ba, Ra
Pd*	Pd, Pt, Au
Be*	Be, Zn, Cd, Hg
K*	K, Rb, Cs, Fr
V*	V, Cr, Nb, Mo, Tc
Ni*	Ni, Cu, Ge, Ru, Rh, Ag, Bi
In*	In, La, Ce, Pr, Nd, Pm, Sm, Eu
Al*	Al, Ga, Y, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Tl
R*	R, He, Ne, Ar, Kr, Xe, Rn, Ac, Th, Pa, U, Np, Pu, Am, Cm, Bk, Cf, Es, Fm, Md, No, Lr, Db, Jl

Table 2. Representation of Phenol by MNA Descriptors of Zero, First, and Second Levels (MNA/0, MNA/1, MNA/2)[a]



atom	MNA/0	MNA/1	MNA/2
1	C	C(CC-H)	C(C(CC-H)C(CC-O)-H(C))
2	C	C(CC-H)	C(C(CC-H)C(CC-H)-H(C))
3	C	C(CC-H)	C(C(CC-H)C(CC-H)-H(C))
4	C	C(CC-H)	C(C(CC-H)C(CC-H)-H(C))
5	C	C(CC-H)	C(C(CC-H)C(CC-O)-H(C))
6	C	C(CC-O)	C(C(CC-H)C(CC-H)-O(C-H))
7	-O	-O(C-H)	-O(C(CC-O)-H(-O))
8	-H	-H(-O)	-H(-O(C-H))
9	-H	-H(C)	-H(C(CC-H))
10	-H	-H(C)	-H(C(CC-H))
11	-H	-H(C)	-H(C(CC-H))
12	-H	-H(C)	-H(C(CC-H))
13	-H	-H(C)	-H(C(CC-H))

[a] Hyphen (-) is the chain marker for the atoms in the chains.

Table 2 shows the structure of phenol presented by MNA descriptors of zero, first, and second levels. For example, for the first atom, C, the zero-level MNA descriptor is "C" (this atom is in cycle), its three neighbors have zero-level MNA descriptors "C", "C", and "-H" (this atom is not in cycle); the first-level MNA descriptor is "C(CC−H)", etc.

In general, at some time of the iteration process the MNA descriptor may cover the molecule completely. However, our experiments have shown that the utilization of MNA descriptors of the first and second levels provides the best accuracy for property prediction.[16]

Such MNA descriptors are generated for each structure from the data set. Each particular descriptor has a unique integer number according to the descriptors' dictionary.

Calculation of Similarity. We have modified the Tanimoto coefficient to take into account the different frequencies of descriptors. The similarity between two molecules, A and B, is given by

$$\mathrm{sim}(A, B) = \frac{\sum_{i=1}^{M} \min[A(i), B(i)]}{\sum_{i=1}^{M} A(i) + \sum_{i=1}^{M} B(i) - \sum_{i=1}^{M} \min[A(i), B(i)]} \quad (1)$$

where A(i) and B(i) are the counts of descriptor i in the molecules A and B, respectively; M is the total number of various descriptors in the dictionary.

Data Sets and Methods of Similarity Assessment Used for Evaluation of MNA Descriptors. Predictive accuracy of MNA descriptors is evaluated on the basis of two sets of chemical compounds (139 hydrocarbons and 15 nitrosamines). These sets have been used before for comparative analysis of several other methods of similarity assessment,[15,16] and therefore they are feasible for evaluation of MNA descriptors.

Four methods discussed earlier are based on two types of graph invariants: (a) numerical graph invariants or topological indices[15] and (b) subgraph invariants called atom pairs.[5]

The Euclidean distance is used as the similarity measure for three of the methods. Different topological indices serve as structural descriptors: (1) the first j principal components, where j is the number of principal components with an eigenvalue greater than 1 and the resulting components scaled to have a variance of 1 (PC_s), (2) the j topological indices extracted by the VARCLUS procedure[15] (TI_u), and (3) the same j topological indices scaled to have a variance of 1 and mean of 0 (TI_s). All these sets have been obtained from the original 96 topological indices by various techniques for reduction of data dimension.[15]

In the fourth method (AP) of similarity assessment, the atom pairs[5] are used as structural descriptors, and the similarity coefficient is calculated as

$$S_{ij} = 2C/(T_i + T_j)$$

where C is the number of atom pairs common to molecules i and j. T_i and T_j are the total number of atom pairs in molecules i and j, respectively.

These four methods have been detailed in ref 15.

Analogue Selection and Property Prediction. Four topology-based methods mentioned above have been used in the paper[15] to quantify the intermolecular similarity of the compounds. By these similarity techniques the nearest neighbors for each compound from the sets of hydrocarbons and nitrosamines have been determined. The value of property (boiling point or mutagenicity) of the nearest neighbor is assigned to the property of the test compound.

Table 3. Hydrocarbons (139) and Their Observed and Predicted Boiling Points[a]

no.	name	obs. bp[15]	pre. MNA	no.	name	obs. bp[15]	pre. MNA
1	*n*-propane	−42.07	−0.50 (2)	71	2,2,3,4-tetramethylpentane	133.02	141.55 (73)
2	*n*-butane	−0.50	36.07 (4)	72	2,2,4,4-tetramethylpentane	122.28	140.27 (70)
3	2-methylpropane	−11.73	27.85 (5)	73	2,3,3,4-tetramethylpentane	141.55	133.02 (71)
4	*n*-pentane	36.07	68.74 (7)	74	benzene	80.10	218.00 (103)
5	2-methylbutane	27.85	−11.73 (3)	75	toluene	110.60	144.40 (77)
6	2,2-dimethylpropane	9.50	0.74 (10)	76	ethylbenzene	136.20	159.20 (80)
7	*n*-hexane	68.74	98.43 (12)	77	*o*-xylene	144.40	139.10 (78)
8	2-methylpentane	60.27	90.05 (13)	78	*m*-xylene	139.10	138.40 (79)
9	3-methylpentane	63.28	91.85 (14)	79	*p*-xylene	138.40	139.10 (78)
10	2,2-dimethylbutane	0.74	79.20 (16)	80	*n*-propylbenzene	159.20	183.30 (87)
11	2,3-dunethylbutane	57.99	89.78 (17)	81	1-methyl-2-ethylbenzene	165.20	161.30 (82)
12	*n*-heptane	98.43	125.67 (21)	82	1-methyl-3-ethylbenzene	161.30	162.00 (83)
13	2-methylhexane	90.05	117.65 (22)	83	1-methyl-4-ethylbenzene	162.00	161.30 (82)
14	3,methylhexane	91.85	118.93 (23)	84	1,2,3-trimethylbenzene	176.10	169.40 (85)
15	3-ethylpentane	93.48	91.85 (14)	85	1,2,4-trimethylbenzene	169.40	176.10 (84)
16	2,2-dimethylpentane	79.20	106.84 (26)	86	1,3,5-trimethylbenzene	164.70	169.40 (85)
17	2,3-dimethylpentane	89.78	115.61 (27)	87	*n*-butylbenzene	183.30	159.20 (80)
18	2,4-dimethylpentane	80.50	109.10 (29)	88	1,2-diethylbenzene	183.40	181.10 (89)
19	3,3-dimethylpentane	86.06	111.97 (30)	89	1,3-diethylbenzene	181.10	183.80 (90)
20	2,2,3-trimethylbutane	80.88	109.84 (34)	90	1,4-diethylbenzene	183.80	181.10 (89)
21	*n*-octane	125.67	150.80 (39)	91	1-methyl-2-*n*-propylbenzene	184.80	181.80 (92)
22	2-methylheptane	117.65	143.26 (40)	92	1-methyl-3-*n*-propylbenzene	181.80	183.80 (93)
23	3-methylheptane	118.93	117.71 (24)	93	1-methyl-4-*n*-propylbenzene	183.80	181.80 (92)
24	4-methylheptane	117.71	118.93 (23)	94	1,2-dimethyl-3-ethylbenzene	193.90	190.00 (96)
25	3-ethylhexane	118.53	141.20 (44)	95	1,2-dimethyl-4-ethylbenzene	189.80	190.00 (96)
26	2,2-dimethylhexane	106.84	132.69 (45)	96	1,3-dimethyl-2-ethylbenzene	190.00	193.90 (94)
27	2,3-dimethylhexane	115.61	140.50 (46)	97	1,3-dimethyl-4-ethylbenzene	188.40	186.90 (99)
28	2,4-dimethylhexane	109.43	133.50 (47)	98	1,3-dimethyl-5-ethylbenzene	183.80	189.80 (95)
29	2,5-dimethylhexane	109.10	135.21 (49)	99	1,4-dimethyl-2-ethylbenzene	186.90	188.40 (97)
30	3,3-dimethylhexane	111.97	137.30 (50)	100	1,2,3,4-tetramethylbenzene	205.00	198.20 (101)
31	3,4-dimethylhexane	117.73	115.65 (32)	101	1,2,3,5-tetramethylbenzene	198.20	196.80 (102)
32	2-methyl-3-ethylpentane	115.65	117.73 (31)	102	1,2,4,5-tetramethylbenzene	196.80	198.20 (101)
33	3-methyl−3-ethylpentane	118.26	140.60 (56)	103	naphthalene	218.00	270.00 (104)
34	2,2,3-trimethylpentane	109.84	133.60 (58)	104	acenaphthalene	270.00	218.00 (103)
35	2,2,4-trimethylpcntane	99.24	124.08 (60)	105	acenaphthene	279.00	359.00 (109)
36	2,3,3-trimethylpentane	114.76	137.68 (61)	106	fluorene	294.00	398.00 (113)
37	2,3,4-trimethylpentane	113.47	139.00 (62)	107	phenanthrene	338.00	383.00 (110)
38	2,2,3,3-tetramethylbutane	106.47	140.27 (70)	108	anthracene	340.00	338.00 (107)
39	*n*-nonane	150.80	125.67 (21)	109	4*H*−cyclopenta(def)phenanthrene	359.00	406.00 (114)
40	2-methyloctane	143.26	144.18 (41)	110	fluoranthene	383.00	338.00 (107)
41	3-methyloctane	144.18	142.48 (42)	111	pyrene	393.00	422.00 (115)
42	4-methyloctane	142.48	144.18 (41)	112	benzo(*a*)fluorene	403.00	406.00 (114)
43	3-elhylheptane	143.00	141.20 (44)	113	benzo(*b*)fluorene	398.00	406.00 (114)
44	4-ethylheptane	141.20	143.00 (43)	114	benzo(*c*)fluorere	406.00	403.00 (112)
45	2,2-dimethylheptane	132.69	137.30 (50)	115	benzo(*ghi*)fluoranthene	422.00	393.00 (111)
46	2,3-dimethylheptane	140.50	133.50 (47)	116	cyclopenta(*cd*)pyrene	439.00	359.00 (109)
47	2,4-dimethylheptane	133.50	136.00 (48)	117	chrysene	431.00	480.00 (122)
48	2,5-dimethylheptane	136.00	133.50 (47)	118	benz(*a*)anthracene	425.00	481.00 (123)
49	2,6- dimethylheptane	135.21	136.00 (48)	119	triphenylene	429.00	481.00 (121)
50	3,3-dimethylheptane	137.30	135.20 (53)	120	naphthacene	440.00	425.00 (118)
51	3,4-dimethylheptane	140.60	138.00 (54)	121	benzo(*b*)fluoranthene	481.00	481.00 (123)
52	3,5- dimethylheptane	136.00	133.80 (55)	122	benzo(*j*)fluoranthene	480.00	493.00 (125)
53	4,4-dimethylheptane	135.20	137.30 (50)	123	benzo(*k*)fluoranthene	481.00	481.00 (121)
54	2-methyl-3-ethylhexane	138.00	140.60 (51)	124	benzo(*a*)pyrene	496.00	534.00 (130)
55	2-methyl-4-ethylhexane	133.80	136.00 (52)	125	benzo(*e*)pyrene	493.00	480.00 (122)
56	3-methyl-3-ethylhexane	140.60	137.30 (50)	126	perylene	497.00	480.00 (122)
57	3-methyl-4-ethylhexane	140.40	140.60 (51)	127	anthanthrene	547.00	542.00 (128)
58	2,2,3-trimethylhexane	133.60	126.54 (59)	128	benzo(*ghi*)perylene	542.00	547.00 (127)
59	2,2,4-trimethylhexane	126.54	133.60 (58)	129	indeno(l,2,3-*cd*)fluoranthene	531.00	480.00 (122)
60	2,2,5-trimethylhexane	124.08	133.60 (58)	130	indeno(1,2,3-*cd*)pyrene	534.00	480.00 (122)
61	2,3,3-trimethylhexane	137.68	133.60 (58)	131	dibenz(*a*,*c*)anthracene	535.00	592.00 (136)
62	2,3,4-trimethylhexane	139.00	136.73 (69)	132	dibenz(*a*,*h*)anthracene	535.00	531.00 (133)
63	2,3,5-trimethylhexane	131.34	139.00 (62)	133	dibenz(*a*,*j*)anthracene	531.00	535.00 (132)
64	2,4,4-trimethylhexane	130.65	126.54 (59)	134	picene	519.00	592.00 (136)
65	3,3,4-trimethylhexane	140.46	133.60 (58)	135	coronene	590.00	547.00 (127)
66	3,3-diethylpentane	146.17	140.60 (56)	136	dibenzo(*a*,*e*)pyrene	592.00	595.00 (139)
67	2,2-dimethyl-3-ethylpentane	133.83	133.60 (58)	137	dibenzo(*a*,*h*)pyrene	596.00	594.00 (138)
68	2,3-dimethyl-3-ethylpentane	142.00	137.68 (61)	138	dibenzo(*a*,*i*)pyrene,	594.00	596.00 (137)
69	2,4-dimethyl-3-ethylpentane	136.73	139.00 (62)	139	dibenzo(*a*,*l*)pyrene	595.00	592.00 (136)
70	2,2,3,3-tetramethylpentane	140.27	122.28 (72)				

[a] The number in parentheses is the number of the nearest neighboring molecule.

Table 4. Comparison of Five Similarity Methods To Select Analogues for Prediction of Boiling Point for 139 Hydrocarbons

similarity method	*r*	SE	*p*
TI_s (15)	0.993	19.34	<0.0001
AP (15)	0.986	26.26	<0.0001
MNA	0.985	27.89	<0.0001
TI_u (15)	0.983	29.36	<0.0001
PC_s (15)	0.966	41.29	<0.0001

Table 5. Nitrosamines and Their Observed and Predicted Mutagenic Potencies

no.	name	ln R^a	pre. ln *R* (MNA)
I	dipropyl-*N*-nitrosamine	−2.53	−1.90 (2)
2	dibutyl-*N*-nitrosamine	−1.90	−3.00 (3)
3	dipentyl-*N*-nitrosamine	−3.00	−1.90 (2)
4	*N*-nitrosopyrrolidine	−3.91	−4.60 (6)
5	*N*-nitrosomorpholine	−2.81	−3.91 (4)
6	*N*-nitrosopiperidine	−4.60	−3.91 (4)
7	*N*-methyl-*N*-nitroso-*N*′-nitroguanidine	7.23	5.86 (8)
8	*N*-ethyl-*N*-nitroso-*N*′-nitroguanidine	5.86	3.69 (9)
9	*N*-propyl-*N*-nitroso-*N*′-nitroguanidine	3.69	3.89 (10)
10	*N*-butyl-*N*-nitroso-*N*′-nitroguanidine	3.89	3.09 (12)
11	*N*-isobutyl-*N*-nitroso-*N*′-nitroguanidine	4.34	5.86 (8)
12	*N*-pentyl-*N*-nitroso-*N*′-nitroguanidine	3.09	1.67 (13)
13	*N*-hexyl-*N*-nitroso-*N*′-nitroguanidine	1.67	3.09 (12)
14	*N*-nitrosomethylurea	1.48	0.10 (15)
15	*N*-nitrosoethylurea	0.10	1.48 (14)

[a] ln *R* represents the natural logarithm of revertants per nanomole (*R*) by the Ames' test. The number in parentheses is the number of the nearest neighboring molecule.

Correlation coefficients and standard errors of the estimates are used to evaluate the relative accuracy of these similarity methods.

Using the same procedure on the basis of MNA descriptors we have estimated the boiling point of hydrocarbons and mutagenicity of nitrosamines from the same sets.

RESULTS

Prediction of the Boiling Point. We calculate the similarity coefficients (eq 1) for all pairs of 139 hydrocarbons using MNA descriptors of the first and second levels. The estimate of the boiling point for each compound is the value of boiling point known for the nearest neighbor molecule. In Table 3 are listed the experimental and calculated values of boiling points for 139 hydrocarbons. The number in parentheses following each predicted boiling point is the number of the nearest neighboring analogue.

Table 4 represents the comparison of our results with those available in the ref 15.

The method of similarity assessment based on MNA descriptors demonstrates highly significant correlation between the observed and predicted boiling points of hydrocarbons ($p < 0.0001$). With this method the boiling point is predicted more precisely than with PC_s and TI_u methods, but less accurately than with AP and TI_s approaches.

Mutagenicity Prediction. The same procedure is used to calculate the similarity coefficients for all pairs of nitrosamines. The mutagenicity of each compound was estimated on the basis of this property for the nearest neighboring molecule. Table 5 represents the experimental mutagenic potencies (ln *R*, natural logarithm of the number of revertants per nanomole) for 15 nitrosamines and mutagenic potencies predicted with the method based on MNA descriptors. Table 6 summarizes the accuracy of each similarity method in mutagenic potency prediction: (1) MNA, (2) PC_s, (3) TI_u, (4) Ti_s, (5) AP, and (6) TEI. The last method is based on topoelectric description of the molecule we have suggested before.[16] It is clear that the most accurate results are obtained by the approach based on MNA descriptors.

Table 6. Comparison of Six Similarity Methods To Select Analogues for Prediction of Mutagenic Potency for 15 Nitrosamines

similarity method	*r*	SE	*p*
MNA	0.945	1.25	<0.0001
AP (15)	0.944	1.26	<0.0001
TEI (16)	0.931	1.43	<0.0001
TI_u (15)	0.923	1.47	<0.0001
PC_s (15)	0.830	2.33	<0.0001
Ti_s (15)	0.740	2.67	<0.0016

DISCUSSION

In this paper we discuss a new approach to the assessment of molecular similarity based on original MNA descriptors. We compare this approach with several other methods for prediction of the boiling point and mutagenicity in the same sets of compounds.[15,16]

The results presented in Tables 4 and 6 show that the approach based on MNA descriptors provides the reasonable estimates for these properties. The expected value of the correlation coefficient for prediction of any property by random selection of the "analogue" from the set of compounds and assigning the apropriate value to the test compound equals zero. However, with the MNA method the correlation coefficients are close to or better than those obtained by the other methods.[15,16]

With this approach the accuracy of prediction of the boiling point is slightly poorer than by TI_s and AP methods. The correlation coefficient *r* and standard error (SE) values for MNA, AP, and TI_s methods are equal to 0.985, 27.89; 0.993, 19.34; and 0.986, 26.26; respectively (Table 4).

With the method based on MNA descriptors, the accuracy of prediction of mutagenicity is close to or better than the best results provided by the AP approach. The *r* and SE values for MNA and AP methods are equal to 0.945, 1.25; and 0.944, 1.26; respectively (Table 6).

The results demonstrate that MNA descriptors can be effectively used at similarity calculations to estimate quite different quantitative properties such as boiling point and mutagenicity. Therefore, MNA descriptors can be applied for both SAR (see refs 9−14) and QSAR/QSPR (this paper) analyses of relationships between the structure and property of chemical compounds.

REFERENCES AND NOTES

(1) Willett, P. Computational tools for the analysis of molecular diversity. *Perspect. Drug Discovery Des.* **1997**, *7/8*, 1−11.
(2) Warr, W. A. Commercial software for diversity analysis. *Perspect. Drug Discovery Des.* **1997**, *7/8*, 115−130.
(3) *Molecular Similarity in Drug Design*; Dean, P. M., Ed.; Chapman & Hall: London, 1995.
(4) Willett, P. *Similarity and Clustering in Chemical Information Systems;* Research studies Press: Letchworth, 1987.
(5) Sheridan, R. P.; Miller, M. D.; Underwood, D. J.; Kearsley, S. K. Chemical Similarity Using Geometric Atom Pair Descriptors. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 128−136.

(6) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atoms Pairs as Molecular Features in Structure−Activity Studies: Definition and Applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64−73.
(7) 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*, 17−39.
(8) Brown, R. D.; Martin, Y. C. Use of Structure−Activity Data to Compare Structure-Based Clustering Methods and Descriptors for Use in Compound selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572−584.
(9) Filimonov, D. A.; Poroikov, V. V.; Karaicheva, E. I.; Kazaryan, R. K.; Boudunova, A. P.; Mikhailovskiy, E. M.; Rudnitskih, A. V.; Goncharenko, L. V.; Burov, Yu. V. Computer-Aided Prediction of Biological Activity Spectra of Chemical Substances on the Basis of Their Structural Formulae: Computerized System PASS. *Exp. Clin. Pharmacol. (Rus)*. **1995**, *58* (2), 56−62.
(10) Filimonov, D. A.; Poroikov, V. V. PASS: Computerized Prediction of Biological Activity Spectra for Chemical Substances. In *Bioactive Compound Design: Possibilities for Industrial Use*; BIOS Scientific Publishers: Oxford, U.K. 1996; pp 47−56.
(11) Poroikov, V. V.; Filimonov, D. A.; Stepanchikova, A. V.; Boudunova, A. P.; Shilova, E. V.; Rudnitskih, A. V.; Goncharenko, L. V. Opimization of Synthesis and Pharmacological Testing of New Compounds Based on Computerized Prediction of Their Biological Activity Spectra. *Chim.-Pharm. J. (Rus)***. 1996**, *30* (9), 20−23. (English translation by Consultants Bureau, New York: *Pharm. Chem. J*. **1996**, *30* (9), 570−573).
(12) Poroikov, V. V.; Filimonov, D. A.; Gloriozova, T. A.; Lagunin, A. A.; Stepanchikova, A. A. Prediction of Biological Activity Spectra for Substances: in House Applications and Internet Feasibility. *2nd International Electronic Conference on Synthetic Organic Chemistry (ECSOC-2)*, http://www.mdpi.org/ecsoc/, September 1−30, 1998, e0004.
(13) http://www.ibmh.msk.su/PASS/.
(14) Gloriozova, T. A.; Filimonov, D. A.; Lagunin, A. A.; Poroikov, V. V. Evaluation of computer system for prediction of biological activity PASS on the set of new chemical compounds. *Chim.-Pharm. J. (Rus).* **1998**, *32* (12), 32−39.
(15) Basak, S. C.; Grunwald, G. D. Molecular Similarity and Risk Assessment: Analogue Selection and Property Estimation Using Graph Invariants. *SAR QSAR Eviron. Res.* **1994**, *2*, 289−307.
(16) Borodina, Yu. V.; Filimonov, D. A.; Poroikov, V. V. Computer-aided estimation of synthetic compounds similarity with endogenous bio-regulators, *Quant. Struct.-Act. Relat.* **1998**, *17*, 459−464.
(17) Cederbaum, L. S. Green's Function for Molecules. *Few Body Systems* **1992**, *Suppl. 6*, 595.

CI980335O