JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

J. Agric. Food Chem. 2010, 58, 2673–2684 2673
DOI:10.1021/jf902639x

Screening Rules for Leads of Fungicides, Herbicides, and Insecticides†

Bin Liu,§ Fucheng Zhu,§ Ying Huang,§ Yuhui Wang,§ Fei Yu,§ Botao Fan,# and Jianhua Yao*,§

§Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Road, Shanghai, China, and #ITODYS, University of Paris 7, Bâtiment Lavoisier 15, rue Jean Antoine de Baïf, 75205 Paris Cedex 13, France

To increase efficiency of finding leads in pesticide design, reasonable screening rules for leads of fungicide, herbicide, and insecticide, respectively, are desired. Previous works showed that "Rule 5" of Lipinski is not a suitable screening rule for leads of pesticide and proposed rules for leads of fungicide, insecticide, and herbicide, which were combined by logarithmic ratio of octanol−water partition coefficient (log *P*), number of hydrogen bond donors, molecular weight, number of hydrogen bond acceptors, polar surface area, carcinogenic toxicity, and mutagenic toxicity. Herein, three sets of screening rules for leads of fungicide, insecticide, and herbicide, respectively, are presented. Each set of screening rules involves seven descriptors, which were selected by Kolmogorov−Smirnov test, ANOVA, Kruskal−Wallis test, and Pearson product-moment correlation, from more than 450 descriptors calculated by Codessa. Their accuracies are about 82, 83, and 89%, respectively.

KEYWORDS: Screening rules; fungicide; insecticide; herbicide; descriptors

INTRODUCTION

Pesticides are applied in farming more seriously because the protection of the environment and human health is paid more attention. At present, the standards for evaluating satisfactory pesticides are maximum effects, minimum toxicity, and environmental friendliness. Drawbacks of traditional approaches used in pesticide design in cost and pollution had appeared. Now, an in silico approach is widely applied in pesticide design because of its advantages. Screening rule is the key of in silico screening. Reasonable screening rules are necessary and important to estimate the potential of compounds to become leads of pesticides.

There were a few published works about screening rules for leads of pesticides. Tice confirmed that Lipinski's "rule of 5" is not fit to determine whether a compound is a possible lead of agrochemicals and proposed screening rules for insecticides and pre-emergence and postemergence herbicides, respectively, which included the molecular properties molecular mass, log *P*, hydrogen bond donor, hydrogen bond acceptor, polar surface area (PSA), and rotatable bonds (*1*, *2*). Our previous work proposed a set of screening rules combining the molecular properties log *P*, molecular weight (MW), number of hydrogen bond acceptors (NHA), number of hydrogen bond donors (NHD), PSA, mutagenic toxicity, and carcinogenic toxicity for leads of herbicides, fungicides, and insecticides, respectively (*3*). Unfortunately, their accuracies are not satisfactory in estimating leads of herbicides, fungicides, or insecticides. Therefore, to find more accurate screening rules, more molecular properties or descriptors should be calculated and analyzed.

In this work, more than 1000 compounds that had one of three bioactivities, fungicide, herbicide, or insecticide, were investigated. More than 450 descriptors of every compound were calculated by Codessa (Comprehensive Descriptors for Structural and Statistical Analysis, v2.63) (*4*). The descriptors are classified into six types: constitutional, that is, relative number of single bonds, gravitation index, etc.; topological, that is, average bonding information content (order 1), Kier and Hall index (order 2), Randic index (order 0), etc.; geometrical, that is, XY shadow, ZX shadow, etc.; electrostatic, that is, atomic charge weighted PPSA, atomic charge weighted PNSA, etc.; quantum-chemical, that is, total charge weighted PPSA, fractional PPSA (PPSA-2/TMSA), etc.; thermodynamics, that is, final heat of formation, internal entropy (300 K), and so on (*5*). The distributions of the descriptors were analyzed by using the Kolmogorov−Smirnov test (*6*), ANOVA (*7*, *8*), and the Kruskal−Wallis test (*9*, *10*). Their correlation is calculated by Pearson product-moment correlation (*11*, *12*).

Herein, we will propose three sets of screening rules for leads of fungicides, herbicides, and insecticides, respectively, which combined seven irredundant descriptors except for the properties mentioned in previous works. They are a complement of the previous screening rules (*1*−*3*).

MATERIALS AND METHODS

Data. In this work, 1285 agrochemicals were selected from various reference sources including the *e-Pesticide Manual* (*13*), a book,

†Part of the ECUST-Qian Pesticide Cluster.

*Corresponding author (telephone +86-21-54925266; fax +86-21-54925264; e-mail yaojh@mail.sioc.ac.cn).

Published on Web 12/10/2009

Table 1. 100 Compounds in the Fungicide Training Set

training set ID	compound	training set ID	compound
2	etaconazole	113	bupirimate
3	tecoram	114	thiophanate
5	zarilamid	117	cypendazole
6	trichlamide	119	kasugamycin
9	2,4-dinitro-6-(1-propylpentyl)phenyl methyl carbonate	124	ofurace
11	fenpropimorph	125	kresoxim-methyl
12	tebuconazole	126	captafol
13	2-heptadecyl-4,5-dihydro-1*H*-imidazole-1-ethanol	128	mecarbinzid
15	quintozene	130	oxycarboxin
17	thioquinox	131	cymoxanil
18	penconazole	136	fludioxonil
20	azaconazole	138	cyproconazole
22	cyprodinil	139	famoxadone
23	milneb	142	iminoctadine
24	anilazine	143	falimorph
25	drazoxolon	144	tridemorph (mixture)
28	thiochlorfenphim	154	spiroxamine
30	mucochloric anhydride	155	thiram
34	fluazinam	159	thiabendazole
36	debacarb	165	ethaboxam
39	natamycin	170	nuarimol
40	ethoxyquin	171	pefurazoate
42	furalaxyl	172	iprovalicarb
43	furametpyr	180	rabenzazole
44	thicyofen	182	tioxymid
45	cyprofuram	183	buthiobate
52	fuberidazole	185	fenaminosulf
57	captan	186	2-(1-ethylhexyl)-4,6-dinitrophenyl methyl carbonate
58	benalaxyl	188	sultropen
66	dichlofluanid	190	2-butanamine
68	azithiram	191	diclobutrazol
69	dichlone	192	isovaledione
70	2,3-dihydro-5-phenyl-1,4-dithi-ine 1,1,4,4-tetraoxide	194	bismerthiazol
74	furcarbanil	199	penthiopyrad
78	2-acetyl-5-methyl-3-oxopent-4-en-5-olide	200	boscalid
82	nitrothal-isopropyl	202	mandipropamid
83	salicylanilide	204	2,3-dichloro-4-(propylsulfonyl)pyridine
84	benodanil	207	2-((hydroxymethyl)amino)-2-methylpropanol
89	ferimzone	209	irgarol
90	oxadixyl	211	tetrachloro-4-(methylsulfonyl)pyridine
97	triazbutil	214	4,6-dichloro-2-phenylphenol
100	carbamorph	217	benzisothizaolin-3-one
101	iprodione	218	dehydroabietylamine
103	halacrinate	219	bis(propylsulfonyl)ethylene
104	imibenconazole	223	2-naphthol
105	hymexazol	227	2,4-dimethyphenol
106	iprobenfos	229	4-nitrophenol
109	phosdiphen	233	4,5-dibromosalicylanilide
111	dodine	234	potassium *N*-hydroxymethyl-*N*-methyldithiocarbamate
112	tetraconazole	235	myclobutanil

Pesticides, An International Guide to 1800 Pest Control Chemicals (*14*), published works (*15*, *16*), U.S. EPA (*17*), and SciFinder (a Chemical Abstracts Service database) (*18*), in which 301 compounds showed fungicide activity, 426 showed insecticide activity, and 558 showed herbicide activity.

Each type of chemical was divided randomly into two sets: training and test. The ratio of training to test was about 5:1. Actually, 253, 355, and 465 compounds were in the fungicide, insecticide, and herbicide training sets, respectively. Some of them, 100 compounds, in each training set are listed in **Tables 1**, **2**, and **3**, respectively, and 48, 71, and 93 compounds in the fungicide, insecticide, and herbicide test sets are listed in **Tables 4**, **5**, and **6**, respectively.

Method. All related descriptors in this paper were calculated by using a package of chemical descriptors, Codessa, which can calculate six types of descriptors: constitutional, topological, geometrical, electrostatic, quantum-chemical, and thermodynamics. Constitutional descriptors are simple descriptors that reflect only the molecular composition of the compound, that is, number of atoms, absolute and relative numbers of C, H, O, S, N, F, Cl, Br, I, and P atoms, number of bonds, number of rings, etc. Topological descriptors (also called topological indices), that is, Wiener index (*19*), information content index, and its derivatives (orders 0−2) (*20*), etc., describe the atomic connectivity in the molecule (*21*, *22*). Geometrical descriptors are concerned with the size, shape, relative position, and properties of space of molecules, that is, shadow indices (*23*), molecular surface area (*24*), and so on. Electrostatic descriptors reflect characteristics of the charge distribution of the molecule, that is, topological electronic index (*25*), charged partial surface area (*26*), etc. Quantum-chemical descriptors are calculated from quantum chemical data (*27*). There are five types of quantum-chemical descriptors: charge distribution-related, valency-related, quantum mechanical energy-related, quantum mechanical molecular rotational−vibrational, and molecular salvation (*5*, *28*). Thermodynamic descriptors relate to molecular interrelation of energy with chemical reactions or with a physical change of state and involve descriptors as follows: thermodynamic heat of formation of the molecule at 300 K, vibrational enthalpy of the molecule (at $T = 300$ K), translational enthalpy of the molecule (at $T = 300$ K), vibrational entropy of the molecule (at $T = 300$ K), rotational entropy of the molecule (at $T = 300$ K), etc. (*5*).

Three sets of descriptors to become the screening rules for leads of fungicides, insecticides, and herbicides, respectively, would be proposed when the analysis work was finished. Herein, four data analysis methods, the K−S test (*6*), ANOVA (*7*, *8*), K−W test (*9*, *10*), and Pearson correlation coefficient (*11*, *12*), were used in the flowchart shown in **Figure 1**.

The Kolmogorov−Smirnov statistic for a given cumulative distribution function $F(x)$ is shown in eq **1**

$$D_n = \mathrm{Sup}[|F_n(x) - F_0(x)|] \quad (1)$$

where $F_n(x)$ is the empirical distribution function of a sample, $F_0(x)$ is a hypothetical distribution function, Sup[ ] is the supremum of set S, n is the sum of samples in the set, and D_n is the Kolmogorov−Smirnov statistic.

ANOVA is a collection of statistical models, and their associated procedures, in which the observed variance is partitioned into components due to different explanatory variables. In principle, there are three models: fixed-effects models, random-effects models, and mixed-effect models. In this work, the fixed-effects models were used.

The K−W test is a nonparametric method for testing the equality of population medians among three or more groups of sample data. It was used in this process if the descriptor was not in normal distribution. The static H is calculated by eq **2**

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \ldots \frac{R_k^2}{n_k}\right) - 3(N+1) \quad (2)$$

where N is the total of samples, k is the total of groups of samples, R_i is the sum of orders of samples in the ith group, and n_i is the sum of samples in the ith group.

The Pearson product-moment correlation coefficient is a measure of the correlation (linear dependence) between two variables X and Y. It is widely used in the sciences as a measure of the strength of linear dependence between two variables. The coefficient (r) between two descriptors in a candidate set was calculated by eq **3**. If r was >0.9 between the two descriptors, the two were correlated, and one of them was removed from the candidate set.

$$r = l_{xy}/\sqrt{l_{xx}l_{yy}} \quad (3)$$

Table 2. 100 Compounds in the Insecticide Training Set

training set ID	compound	training set ID	compound
4	butathiofos	138	azothoate
5	chlorthiophos	144	dimefox
7	dieldrin	146	bufencarb
8	cycloprothrin	148	carbon disulfide
9	biopermethrin; biopermethrine	149	carbophenothion
12	diethyl 5-methylpyrazol-3-yl phosphate	152	chlormephos
17	2-isovalerylindan-1,3-dione	154	chlorprazophos
18	2-methyl(prop-2-ynyl)aminophenyl methylcarbamate	155	butocarboxim
24	azadirachtin	156	cyanthoate
25	amitraz	163	dialifos; dialiphos; dialifor
26	azamethiphos	165	cartap
27	azinphos-methyl	174	carbofuran
29	azinphos-ethyl	183	aldicarb
33	cyfluthrin	185	chlorthiophos
34	allyxycarb (BSI, E-ISO); allyxycarbe; APC	196	cevadine
37	buprofezin	199	chloropicrin
38	4-methyl(prop-2-ynyl)amino-3,5-xylyl methylcarbamate	222	cyromazine
44	dichlorvos	224	chlorfluazuron
46	HHDN	227	chlorfenapyr
53	decarbofuran	228	chromafenozide
55	anabasine	231	cyanophos
57	diazinon	234	chlorpyrifos
64	dichloroethyl ether	236	DDT
65	bioresmethrin	243	chlorpyrifos-methyl
68	acephate	244	bendiocarb
71	chlorbicyclen; chlorbicyclene	245	cadusafos
72	carbosulfan	251	alanycarb
73	diflubenzuron	266	aldoxycarb; aldoxycarbe
75	butoxycarboxim	270	carbanolate
76	diafenthiuron	278	2-(2-butoxyethoxy)ethyl thiocyanate
80	1-bromo-2-chloroethane	280	acrylonitrile
81	butonate	282	Bayer 22/190
82	2,2-dichlorovinyl 2-ethylsulfinylethyl methyl phosphate	298	(*Z*)-dodec-8-en-1-ol
83	deltamethrin	300	(*S*)-methoprene
86	dicrotophos	302	methyl isophenphos
87	dicyclanil	305	dimethenamid-P
89	2-(4,5-dimethyl-1,3-dioxolan-2-yl)phenyl methylcarbamate	310	(11*Z*,13*Z*)-hexadecadienal
90	2-(4-chloro-3,5-xylyloxy)ethanol	313	(*Z*,*Z*)-3,13-octadecadien-1-ol acetate
93	Bayer 22408	314	2,6-dinitro-4-octylphenyl crotonate
94	bromfenvinfos	315	3-hydroxy-1-octene
95	bioallethrin	316	3-methyl-2-cyclohexenone
97	2-chlorovinyl diethyl phosphate	319	1,2,4-trimethoxybenzene
99	athidathion	320	DEET
100	amidithion; amidiphos	321	carbofuran phenol
101	chlorfenvinphos	324	7,11-hexadecadien-1-ol, acetate, (7*Z*,11*E*)-
115	3-bromo-1-chloroprop-1-ene	326	(*E*)-*N*-((4-hydroxy-3-methoxyphenyl)-methyl)-8-methyl-6-nonenamide
118	3-methyl-1-phenylpyrazol-5-yl dimethylcarbamate	329	bomyl
121	5,5-dimethyl-3-oxocyclohex-1-enyl dimethylcarbamate	335	allantoin
130	2-(4-chloro-3,5-xylyloxy)ethanol	349	chlorphoxim
132	DAEP	352	bistrifluoron

l_{xx} is calculated by eq **4**, and l_{yy} is calculated by eq **5**.

$$l_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1) \quad (4)$$

$$l_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})/(n-1) \quad (5)$$

The corresponding four modules in SPSS version 13.0 (*29*) were used to do the K−S test, ANOVA, K−W test, and Pearson correlation. In the analysis processing, CISOC-LSS (*30*) was used to get scatter plots and bar charts of analysis objects.

RESULTS AND DISCUSSION

Six types of descriptors, constitutional, topological, geometrical, electrostatic, quantum-chemical, and thermodynamic, including 485 descriptors, were calculated for each compound in the three training sets by Codessa. Among these descriptors, 155 descriptors were element-dependent, 189 were element-independent and in continuous distribution, and 34 were element-independent and in discrete distribution; 107 descriptors were element-independent and equal to zero. The 189 descriptors listed in **Table 7**, selected from the 485 descriptors, were analyzed further by K−S test, ANOVA, K−W test, and scatter plots.

The K−S test was used to analyze the 189 descriptors. The corresponding result showed that 17 descriptors (listed in **Table 8**) were in normal population and 172 were not in normal population. ANOVA was used to calculate differences among the 17 descriptors (listed in **Table 8**). Except for the geometrical descriptor, XY shadow/XY rectangle, which could not be a candidate for screening rules because it had no significant difference among the three types of compounds in the training sets, the

Table 3. 100 Compounds in the Herbicide Training Set

training set ID	compound	training set ID	compound
4	2,4-D	132	bromobonil
9	pentanochlor	134	dinosam
10	ethofumesate	143	cliodinate
12	2,4,5-T	145	simetryn
15	dichlormate	148	thifensulfuron
16	carfentrazone-ethyl	159	chloroxynil
17	trifop	160	benzipram
18	trifop-methyl (unsaturated stereochemistry)	162	trimeturon
19	sulfentrazone	165	fluothiuron
20	chlorimuron-ethyl	189	amiprofos-methyl
22	tri-allate	210	EXD
23	trietazine	211	dimexano
24	UBI-S734	244	chloroacetic acid
27	dimethametryn	245	chlorimuron
28	daimuron	246	fluoroxypyr-meptyl
29	tricamba	247	fluoroxypyr-2-butoxy-1-methylethyl
33	karbutilate	248	flamprop-M
34	azafenidin	249	flamprop-M-methyl
36	azimsulfuron	273	imazapyr
37	prometryn	282	isoxaben
38	amitrole	285	chlorflurenol-methyl
39	2,4,5-TB	299	sulfosulfuron
45	diflufenican	300	2,4-D-butotyl
46	benfluralin	301	tepraloxydim
48	isopropalin	302	monolinuron
49	butachlor	303	cinmethylin
52	aclonifen	304	anilofos
53	dicamba	305	molinate
54	halosafen	306	asulam
55	isocil	307	atrazine
58	amidosulfuron	309	benazolin
67	fluoroxypyr	310	2,4-D-butyl
68	MCPA	311	imazamethabenz
69	metazachlor	314	benazolin-ethyl
70	bromoxynil	320	ethalfluralin
72	propyzamide	337	metolachlor
73	propaquizafop	362	procyazine
75	metamitron	363	nitrofluorfen
76	flamprop-M-isopropyl	364	prynachlor
79	2,4-D-isoctyl	371	methoxyphenone
83	bentazone	372	diethatyl-ethyl
84	fluchloralin	389	di-allate
86	isopolinate	392	cyprazole
91	morfamquat dichloride	402	amibuzin
94	1,1,1,3,3,3-hexafluoro-2-propanone	412	benzoylprop
96	perfluidone	413	2,4,5-T-isoctyl
99	tetrafluoron	423	pyrithiobac
100	methiuron	433	isoxachlortole
124	pyriclor	436	2,4-DB, butoxyethyl ester
131	terbucarb	439	2-(M-chlorophenoxy)-propionamide

others became candidates. The K−W test was used to handle the 172 descriptors not in normal population. The result showed that 34 descriptors (in **Table 9**) were not candidates because they had no significant difference (significance ≥ 0.05) among the three types of compounds. Therefore, after the K−S test, ANOVA, and K−W test, we got a set of candidates including 154 descriptors. When scatter plots and bar charts of the 154 descriptors were analyzed, 35 descriptors (listed in **Table 10**) showed characters of screening rules for leads of fungicides, insecticides, and herbicides, respectively. Corresponding data ranges are shown in **Table 11**.

Table 4. 48 Compounds in the Fungicide Test Set

test set ID	compound	test set ID	compound
1	fenapanil	25	etridiazole
2	4-(1-ethylhexyl)-2,6-dinitrophenyl methyl carbonate	26	tolylfluanid
3	binapacryl	27	imazalil
4	ICIA 0858	28	fenamidone
5	azoxystrobin	29	flusulfamide
6	etem	30	chlorothalonil
7	(*RS*)-*N*-(3,5-dichlorophenyl)-2-(methoxymethyl)succinimide	31	dinobuton
8	SSF109 (shionogi)	32	ethirimol
9	benzamorf	33	flutolanil
10	chloroneb	34	metazoxolon
11	prochloraz	35	mebenil
12	2-pyridinethiol 1-oxide	36	ditalimfos
13	methfuroxam	37	polyoxin D
14	thifluzamide	38	BR enantiomer
15	biphenyl	39	proquinazid
16	hexachlorobenzene	40	chloro-2-cyclopentylphenol
17	dichlozoline	41	5-chlorosalicylanilide
18	diphenylamine	42	(1,1′-biphenyl)-2-ol, 5-chloro-
19	dicloran	43	glutaraldehyde
20	bromuconazole	44	4-*tert*-butylphenol,
21	hexaconazole	45	aureonuclemycin
22	*sec*-butylamine	46	metominostrobin
23	prothiocarb	47	myclozolin
24	hexylthiofos	48	piperalin

Sixteen groups of descriptors listed in **Table 12** were generated when the 35 descriptors were clustered according to their definition. If one descriptor was selected from each group, then 41472 combinations of candidate sets were generated. Also, Pearson correlation coefficients among the 35 for the three types of compounds, respectively, were calculated and are given in the Supporting Information (STable 1). Three sets with higher sensitivity and low coefficients for leads of fungicides, insecticides, and herbicides, respectively, are shown in **Tables 13**, **14**, and **15** when they were tested by compounds in the three test sets listed in **Tables 4**, **5**, and **6**, respectively.

All descriptors listed in **Tables 13**, **14**, and **15** were clustered again by Pearson correlation coefficients listed in the Supporting Information (STable 1) and were tested by the three test sets.

Fungicide. The descriptors in **Table 13** were grouped into (1), (4, 13, 15, 20, 26, 34), (9), (12), (16, 18), (29), (10, 33), (24), and (25) by Pearson's coefficients. They were tested by the three sets, and the results in **Table 16** show that the two electrostatic descriptors, 24 (PPSA-3 atomic charge weighted PPSA [Zefirov's PC]) and 25 (RPCG relative positive charge (QMPOS/QTPLUS) [Zefirov's PC]), were not able to differentiate the three types of compounds because they did not present the characters of fungicides among the three types of compounds. The two descriptors were removed from candidate sets. The screening rules for leads of fungicides involved seven descriptors, average bonding information content (order 1), FNSA-2 fractional PNSA (PNSA-2/TMSA) [Zefirov's PC], final heat of formation/number of atoms, internal enthalpy (300 K)/number of atoms, Randic index (order 0), total molecular one-center E-N attraction/number of atoms, and WNSA-3 weighted PNSA (PNSA3*TMSA/1000) [quantum-chemical PC], and their data ranges are listed in **Table 17**. They were tested by three test sets, and the best result is listed in **Table 18**.

Table 5. 71 Compounds in the Insecticide Test Set

test set ID	compound	test set ID	compound
1	γ-benzene hexachloride	37	profenofos
2	acrinathrin	38	demeton-S
3	β-cypermethrin	39	trimedlure-A
4	2,2,2-trichloro-1-(3,4-dichlorophenyl)ethyl acetate	40	hydrogen cyanide
5	thiocarboxime	41	clothianidin
6	transpermethrin	42	coumaphos
7	aminocarb	43	avermectin B1b
8	sulfoxide (ESA)	44	bifenthrin
9	benfuracarb	45	chlordane
10	methocrotophos	46	cyphenothrin [(1*R*)-*trans*- isomers]
11	fenobucarb	47	thiofanox
12	1,1-dichloro-2,2-bis(4-ethylphenyl)ethane	48	monocrotophos
13	transfluthrin	49	phenthoate
14	chlordecone	50	mecarbam
15	ethion	51	sulfluramid
16	fosmethilan	52	emamectin
17	2-(1,3-dithiolan-2-yl)phenyl dimethylcarbamate	53	ethohexadiol
18	fosthietan	54	methacrifos
19	hydroprene	55	1,1-dichloro-1-nitroethane
20	ethoate-methyl	56	nitrilacarb
21	2-thiocyanatoethyl laurate	57	phenkapton
22	heptachlor	58	pirimiphos-ethyl
23	coumithoate	59	fenazox
24	DSP	60	nicotine
25	leptophos	61	ethiprole
26	hydramethylnon	62	tetradecanal
27	IPSP	63	oxypurinol
28	methoxychlor	64	*trans*-phosphamidon
29	bromophos-ethyl	65	(*E*)-6-dodecen-1-yl acetate
30	isopropyl *O*-(methoxyamino-thiophosphoryl)salicylate	66	5-pentyl dihydrofuranone
31	cyanofenphos	67	benzocaine
32	demeton-O	68	1*H*-pyrazole-3-carbonitrile
33	dicapthon	69	novaluron
34	morphothion	70	methothrin
35	fenitrothion	71	flupyrazofos
36	R-1492		

Table 6. 93 Compounds in the Herbicide Test Set

test set ID	compound	test set ID	compound
1	propachlor	48	pretilachlor
2	chlorfenac	49	diquat dibromide
3	cyanatryn	50	halosulfuron
4	clomazone	51	2,4-DB-butyl
5	dinofenate	52	flucarbazone
6	isoxaben	53	clodinafop
7	phenmedipham-ethyl	54	metribuzin
8	SMY 1500	55	flazasulfuron
9	iodobonil	56	flumioxazin
10	fluometuron	57	fluridone
11	EL 177	58	ethametsulfuron-methyl
12	chlorthiamid	59	haloxyfop
13	FMC 19873	60	simazine
14	pyridate	61	terbutryn
15	trifopsime	62	2,4-D-isopropyl
16	dichlobenil	63	chlorbromuron
17	rimsulfuron	64	cyclosulfamuron
18	sulcotrione	65	dimethenamid
19	ACD 10614	66	dimepiperate
20	bensulfuron-methyl	67	acetochlor
21	fluoronitrofen	68	EPTC
22	buthiuron	69	fosamine-ammonium
23	bensulfuron	70	tribenuron-methyl
24	benzofenap	71	fluthiacet-methyl
25	diclofop	72	tridiphane
26	tralkoxydim	73	dinoseb acetate
27	chlorotoluron	74	dipropetryn
28	sulglycapin	75	LS830556
29	propham	76	phenobenzuron
30	dimethipin	77	flumezin
31	MCPA	78	chlorfenprop (racemate)
32	credazine	79	diethamquat dichloride
33	chloroxuron	80	thidiazimin
34	benzofluor	81	2,4-DEP
35	brompyrazon	82	eglinazine-ethyl
36	fluoromidine	83	proglinazine
37	flumetsulam	84	clodinafop-propargyl
38	diuron	85	pyriminobac-methyl
39	fenoxaprop-P-ethyl	86	profluazol
40	chlorazifop-propargyl (*R*)-isomer	87	4,5-dichloro-2-*N*-octyl-3(2*H*)-isothiazolone
41	butroxydim	88	MCPA, isopropyl ester
42	imazapic	89	picolinafen
43	ethiolate	90	AKH-7088
44	proxan	91	quizalofop-ethyl
45	carbetamide	92	orbencarb
46	pyraflufen-ethyl	93	naptalam
47	mefluidide		

Information in **Table 18** shows that the accuracy of the set of screening rules for leads of fungicides was 82%; those for leads of insecticides and herbicides were 56 and 70%, respectively. It was more suitable for fungicide leads than the others because it only made 18% error for this type of lead, but 44 and 30% errors for the other two types of leads, respectively.

Insecticide. The descriptors in **Table 14** were grouped into (2), (4, 15, 17, 20, 23, 26, 30, 35), (5), (7), (8), (10, 23), (11), (29), and (15, 23, 24, 30) by Pearson's coefficients. They were tested by the three sets, and the results in **Table 19** show that two descriptors, 2 (average information content (order 1), a topological descriptor) and 7 (ESP-RPCG relative positive charge (QMPOS/QTPLUS) [quantum-chemical PC], a quantum-chemical descriptor), were not able to differentiate the three types of compounds because they could not present the characters of insecticides among the three types of compounds. They were removed from candidate sets. The screening rules for leads of insecticides involved seven descriptors, ESP-DPSA-3 difference in CPSAs (PPSA3-PNSA3) [quantum-chemical PC], FNSA-2 fractional PNSA (PNSA-2/TMSA) [quantum-chemical PC], FPSA-2 fractional PPSA (PPSA-2/TMSA) [quantum-chemical PC], final heat of formation, PPSA-3 atomic charge weighted PPSA [Zefirov's PC], total molecular one-center E-N attraction/number of atoms, and ZX shadow, and their data ranges are listed in **Table 20**. They were tested by three test sets, and the best result is listed in **Table 21**.

Information in **Table 21** shows that the accuracy of the set of screening rules for leads of insecticides was 83%. Those for leads of fungicides and herbicides were 60 and 74%, respectively. It was more suitable for insecticide leads than the others because it only made 17% error for this type of lead compounds, but 40 and 26% errors for the other two types of leads, respectively.

Herbicide. The descriptors in **Table 15** were grouped into (1), (7), (13, 14, 15, 23, 24, 26, 28), (9, 22), (10, 23), (16, 29), (16, 32), and (20, 14, 28) by Pearson's coefficients. They were tested by the three sets, and the results in **Table 22** show that a

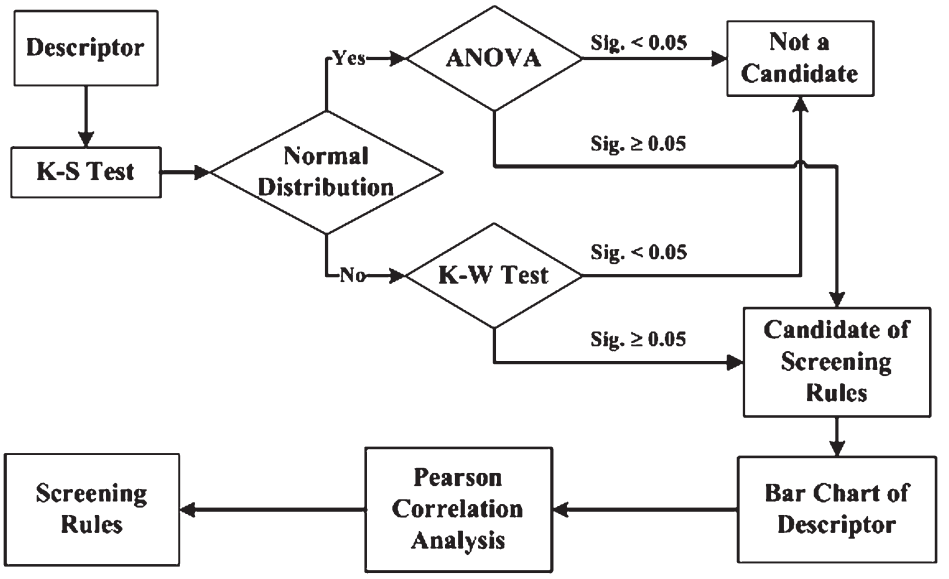

Table 7. 189 Element-Independent Descriptors in Continuous Distribution

no.	descriptor	type
1	(1/2)× BETA polarizability (DIP)	electrostatic
2	(1/6)× GAMMA polarizability (DIP)	electrostatic
3	1× BETA polarizability (DIP)	electrostatic
4	1× GAMMA polarizability (DIP)	electrostatic
5	ALFA polarizability (DIP)	electrostatic
6	average bonding information content (order 0)	topological
7	average bonding information content (order 1)	topological
8	average bonding information content (order 2)	topological
9	average complementary information content (order 0)	topological
10	average complementary information content (order 1)	topological
11	average complementary information content (order 2)	topological
12	average information content (order 0)	topological
13	average information content (order 1)	topological
14	average information content (order 2)	topological
15	average structural information content (order 0)	topological
16	average structural information content (order 1)	topological
17	average structural information content (order 2)	topological
18	Balaban index	topological
19	bonding information content (order 0)	topological
20	bonding information content (order 1)	topological
21	bonding information content (order 2)	topological
22	complementary information content (order 0)	topological
23	complementary information content (order 1)	topological
24	complementary information content (order 2)	topological
25	DPSA-1 difference in CPSAs (PPSA1-PNSA1) [quantum-chemical PC]	quantum-chemical
26	DPSA-1 difference in CPSAs (PPSA1-PNSA1) [Zefirov's PC]	electrostatic
27	DPSA-2 difference in CPSAs (PPSA2-PNSA2) [quantum-chemical PC]	quantum-chemical
28	DPSA-2 difference in CPSAs (PPSA2-PNSA2) [Zefirov's PC]	electrostatic
29	DPSA-3 difference in CPSAs (PPSA3-PNSA3) [Zefirov's PC]	electrostatic
30	ESP-DPSA-1 difference in CPSAs (PPSA1-PNSA1) [quantum-chemical PC]	quantum-chemical
31	ESP-DPSA-2 difference in CPSAs (PPSA2-PNSA2) [quantum-chemical PC]	quantum-chemical
32	ESP-DPSA-3 difference in CPSAs (PPSA3-PNSA3) [quantum-chemical PC]	quantum-chemical
33	ESP-FNSA-1 fractional PNSA (PNSA-1/TMSA) [quantum-chemical PC]	quantum-chemical
34	ESP-FNSA-2 fractional PNSA (PNSA-2/TMSA) [quantum-chemical PC]	quantum-chemical
35	ESP-FNSA-3 fractional PNSA (PNSA-3/TMSA) [quantum-chemical PC]	quantum-chemical
36	ESP-FPSA-1 fractional PPSA (PPSA-1/TMSA) [quantum-chemical PC]	quantum-chemical
37	ESP-FPSA-2 fractional PPSA (PPSA-2/TMSA) [quantum-chemical PC]	quantum-chemical
38	ESP-FPSA-3 fractional PPSA (PPSA-3/TMSA) [quantum-chemical PC]	quantum-chemical
39	ESP-Min net atomic charge	quantum-chemical
40	ESP-PNSA-1 partial negative surface area [quantum-chemical PC]	quantum-chemical
41	ESP-PNSA-2 total charge weighted PNSA [quantum-chemical PC]	quantum-chemical
42	ESP-PNSA-3 atomic charge weighted PNSA [quantum-chemical PC]	quantum-chemical
43	ESP-PPSA-1 partial positive surface area [quantum-chemical PC]	quantum-chemical
44	ESP-PPSA-2 total charge weighted PPSA [quantum-chemical PC]	quantum-chemical
45	ESP-PPSA-3 atomic charge weighted PPSA [quantum-chemical PC]	quantum-chemical

Table 7. Continued

no.	descriptor	type
46	ESP-RNCG relative negative charge (QMNEG/QTMINUS) [quantum-chemical PC]	quantum-chemical
47	ESP-RNCS relative negative charged SA (SAMNEG*RNCG) [quantum-chemical PC]	quantum-chemical
48	ESP-RPCG relative positive charge (QMPOS/QTPLUS) [quantum-chemical PC]	quantum-chemical
49	ESP-RPCS relative positive charged SA (SAMPOS*RPCG) [quantum-chemical PC]	quantum-chemical
50	ESP-TMSA total molecular surface area [quantum-chemical PC]	quantum-chemical
51	ESP-WNSA-1 weighted PNSA (PNSA1*TMSA/1000) [quantum-chemical PC]	quantum-chemical
52	ESP-WNSA-2 weighted PNSA (PNSA2*TMSA/1000) [quantum-chemical PC]	quantum-chemical
53	ESP-WNSA-3 weighted PNSA (PNSA3*TMSA/1000) [quantum-chemical PC]	quantum-chemical
54	ESP-WPSA-1 weighted PPSA (PPSA1*TMSA/1000) [quantum-chemical PC]	quantum-chemical
55	ESP-WPSA-2 weighted PPSA (PPSA2*TMSA/1000) [quantum-chemical PC]	quantum-chemical
56	ESP-WPSA-3 weighted PPSA (PPSA3*TMSA/1000) [quantum-chemical PC]	quantum-chemical
57	final heat of formation	thermodynamic
58	final heat of formation/no. of atoms	thermodynamic
59	FNSA-1 fractional PNSA (PNSA-1/TMSA) [quantum-chemical PC]	quantum-chemical
60	FNSA-1 fractional PNSA (PNSA-1/TMSA) [Zefirov's PC]	electrostatic
61	FNSA-2 fractional PNSA (PNSA-2/TMSA) [quantum-chemical PC]	quantum-chemical
62	FNSA-2 fractional PNSA (PNSA-2/TMSA) [Zefirov's PC]	electrostatic
63	FNSA-3 fractional PNSA (PNSA-3/TMSA) [quantum-chemical PC]	quantum-chemical
64	FNSA-3 fractional PNSA (PNSA-3/TMSA) [Zefirov's PC]	electrostatic
65	FPSA-1 fractional PPSA (PPSA-1/TMSA) [quantum-chemical PC]	quantum-chemical
66	FPSA-1 fractional PPSA (PPSA-1/TMSA) [Zefirov's PC]	electrostatic
67	FPSA-2 fractional PPSA (PPSA-2/TMSA) [quantum-chemical PC]	quantum-chemical
68	FPSA-2 fractional PPSA (PPSA-2/TMSA) [Zefirov's PC]	electrostatic
69	FPSA-3 fractional PPSA (PPSA-3/TMSA) [quantum-chemical PC]	quantum-chemical
70	FPSA-3 fractional PPSA (PPSA-3/TMSA) [Zefirov's PC]	electrostatic
71	gravitation index (all bonds)	constitutional
72	gravitation index (all pairs)	constitutional
73	HOMO − LUMO energy gap	quantum-chemical
74	HOMO energy	quantum-chemical
75	HOMO-1 energy	quantum-chemical
76	omage of the Onsager−Kirkwood solvation energy	quantum-chemical
77	information content (order 0)	topological
78	information content (order 1)	topological
79	information content (order 2)	topological
80	internal enthalpy (300 K)	thermodynamic
81	internal enthalpy (300 K)/no. of atoms	thermodynamic
82	internal entropy (300 K)	thermodynamic
83	internal entropy (300 K)/no. of atoms	thermodynamic
84	internal heat (300 K)	thermodynamic
85	internal heat (300 K)/no. of atoms	thermodynamic
86	Kier flexibility index	topological
87	Kier shape index (order 1)	topological
88	Kier shape index (order 2)	topological
89	Kier shape index (order 3)	topological
90	Kier and Hall index (order 0)	topological
91	Kier and Hall index (order 1)	topological
92	Kier and Hall index (order 2)	topological
93	Kier and Hall index (order 3)	topological
94	lowest normal mode vibrational frequency	quantum-chemical
95	molecular surface area	geometrical
96	molecular volume	geometrical
97	molecular volume/XYZ box	geometrical
98	molecular weight	constitutional
99	moment of inertia A	geometrical
100	moment of inertia B	geometrical
101	moment of inertia C	geometrical
102	no. of occupied electronic levels	quantum-chemical
103	no. of occupied electronic levels/no. of atoms	quantum-chemical
104	PNSA-1 partial negative surface area [quantum-chemical PC]	quantum-chemical
105	PNSA-1 partial negative surface area [Zefirov's PC]	electrostatic
106	PNSA-2 total charge weighted PNSA [quantum-chemical PC]	quantum-chemical
107	PNSA-2 total charge weighted PNSA [Zefirov's PC]	electrostatic
108	PNSA-3 atomic charge weighted PNSA [Zefirov's PC]	electrostatic
109	PPSA-1 partial positive surface area [quantum-chemical PC]	quantum-chemical
110	PPSA-1 partial positive surface area [Zefirov's PC]	electrostatic
111	PPSA-2 total charge weighted PPSA [quantum-chemical PC]	quantum-chemical
112	PPSA-2 total charge weighted PPSA [Zefirov's PC]	electrostatic
113	PPSA-3 atomic charge weighted PPSA [quantum-chemical PC]	quantum-chemical

Table 7. Continued

no.	descriptor	type
114	PPSA-3 atomic charge weighted PPSA [Zefirov's PC]	electrostatic
115	principal moment of inertia A	thermodynamic
116	principal moment of inertia A/no. of atoms	thermodynamic
117	principal moment of inertia B	thermodynamic
118	principal moment of inertia B/no. of atoms	thermodynamic
119	principal moment of inertia C	thermodynamic
120	principal moment of inertia C/no. of atoms	thermodynamic
121	Randic index (order 0)	topological
122	Randic index (order 1)	topological
123	Randic index (order 2)	topological
124	Randic index (order 3)	topological
125	relative molecular weight	constitutional
126	relative no. of single bonds	constitutional
127	RNCG relative negative charge (QMNEG/QTMINUS) [quantum-chemical PC]	quantum-chemical
128	RNCG relative negative charge (QMNEG/QTMINUS) [Zefirov's PC]	electrostatic
129	RNCS relative negative charge SA (SAMNEG*RNCG) [Zefirov's PC]	electrostatic
130	rotational enthalpy (300 K)/no. of atoms	thermodynamic
131	rotational entropy (300 K)	thermodynamic
132	rotational entropy (300 K)/no. of atoms	thermodynamic
133	rotational heat capacity (300 K)/no. of atoms	thermodynamic
134	RPCG relative positive charge (QMPOS/QTPLUS) [Zefirov's PC]	quantum-chemical
135	RPCS relative positive charged SA (SAMPOS*RPCG) [quantum-chemical PC]	quantum-chemical
136	RPCS relative positive charged SA (SAMPOS*RPCG) [Zefirov's PC]	electrostatic
137	structural information content (order 0)	topological
138	structural information content (order 1)	topological
139	structural information content (order 2)	topological
140	TMSA total molecular surface area [quantum-chemical PC]	quantum-chemical
141	TMSA total molecular surface area [Zefirov's PC]	electrostatic
142	topographic electronic index (all bonds) [Zefirov's PC]	electrostatic
143	topographic electronic index (all pairs) [Zefirov's PC]	electrostatic
144	total dipole of the molecule	quantum-chemical
145	total enthalpy (300 K)	quantum-chemical
146	total enthalpy (300 K)/no. of atoms	quantum-chemical
147	total entropy (300 K)	quantum-chemical
148	total entropy (300 K)/no. of atoms	thermodynamic
149	total heat capacity (300 K)	thermodynamic
150	total heat capacity (300 K)/no. of atoms	thermodynamic
151	total hybridization comp of the molecular dipole	quantum-chemical
152	total molecular one-center E-E repulsion	quantum-chemical
153	total molecular one-center E-E repulsion/no. of atoms	quantum-chemical
154	total molecular one-center E-N attraction	quantum-chemical
155	total molecular one-center E-N attraction/no. of atoms	quantum-chemical
156	total molecular two-center resonance energy	quantum-chemical
157	total molecular two-center resonance energy/no. of atoms	quantum-chemical
158	total point-charge comp of the molecular dipole	quantum-chemical
159	translational enthalpy (300 K)/no. of atoms	thermodynamic
160	translational entropy (300 K)	thermodynamic
161	translational entropy (300 K)/no. of atoms	thermodynamic
162	translational heat capacity (300 K)/no. of atoms	thermodynamic
163	vibrational enthalpy (300 K)	thermodynamic
164	vibrational enthalpy (300 K)/no. of atoms	thermodynamic
165	vibrational entropy (300 K)	thermodynamic
166	vibrational entropy (300 K)/no. of atoms	thermodynamic
167	vibrational heat capacity (300 K)	thermodynamic
168	vibrational heat capacity (300 K)/no. of atoms	thermodynamic
169	Wiener index	topological
170	WNSA-1 weighted PNSA (PNSA1*TMSA/1000) [quantum-chemical PC]	quantum-chemical
171	WNSA-1 weighted PNSA (PNSA1*TMSA/1000) [Zefirov's PC]	electrostatic
172	WNSA-2 weighted PNSA (PNSA2*TMSA/1000) [quantum-chemical PC]	quantum-chemical
173	WNSA-2 weighted PNSA (PNSA2*TMSA/1000) [Zefirov's PC]	electrostatic
174	WNSA-3 weighted PNSA (PNSA3*TMSA/1000) [quantum-chemical PC]	quantum-chemical
175	WNSA-3 weighted PNSA (PNSA3*TMSA/1000) [Zefirov's PC]	electrostatic
176	WPSA-1 weighted PPSA (PPSA1*TMSA/1000) [quantum-chemical PC]	quantum-chemical
177	WPSA-1 weighted PPSA (PPSA1*TMSA/1000) [Zefirov's PC]	electrostatic
178	WPSA-2 weighted PPSA (PPSA2*TMSA/1000) [quantum-chemical PC]	quantum-chemical
179	WPSA-2 weighted PPSA (PPSA2*TMSA/1000) [Zefirov's PC]	electrostatic
180	WPSA-3 weighted PPSA (PPSA3*TMSA/1000) [quantum-chemical PC]	quantum-chemical
181	WPSA-3 weighted PPSA (PPSA3*TMSA/1000) [Zefirov's PC]	electrostatic

Table 7. Continued

no.	descriptor	type
182	XY shadow	geometrical
183	XY shadow/XY rectangle	geometrical
184	YZ shadow	geometrical
185	YZ shadow/YZ rectangle	geometrical
186	zero-point vibrational energy	thermodynamic
187	zero-point vibrational energy/no. of atoms	thermodynamic
188	ZX shadow	geometrical
189	ZX shadow/ZX rectangle	geometrical

Table 8. 17 Descriptors in Normal Population

no.	descriptor	no.	descriptor
1	bonding information content (order 0)	10	total entropy (300 K)
2	DPSA-1 difference in CPSAs (PPSA1-PNSA1) [quantum-chemical PC]	11	total molecular two-center resonance energy/no. of atoms
3	DPSA-1 difference in CPSAs (PPSA1-PNSA1) [Zefirov's PC]	12	vibrational enthalpy (300 K)/no. of atoms
4	ESP-DPSA-1 difference in CPSAs (PPSA1-PNSA1) [quantum-chemical PC]	13	vibrational entropy (300 K)
5	ESP-PPSA-1 partial positive surface area [quantum-chemical PC]	14	XY shadow/XY rectangle
6	internal entropy (300 K)	15	YZ shadow
7	molecular surface area	16	YZ shadow/YZ rectangle
8	PPSA-1 partial positive surface area [Zefirov's PC]	17	ZX shadow/ZX rectangle
9	structural information content (order 0)		

Table 9. 34 Descriptors Removed from the Candidate Set by K−W Test

no.	descriptor	significance
1	(1/6)× GAMMA polarizability (DIP)	0.427
2	1× GAMMA polarizability (DIP)	0.427
3	ALFA polarizability (DIP)	0.415
4	average complementary information content (order 0)	0.209
5	complementary information content (order 0)	0.294
6	ESP-FNSA-1 fractional PNSA (PNSA-1/TMSA) [quantum-chemical PC]	0.247
7	ESP-FNSA-3 fractional PNSA (PNSA-3/TMSA) [quantum-chemical PC]	0.442
8	ESP-FPSA-1 fractional PPSA (PPSA-1/TMSA) [quantum-chemical PC]	0.247
9	ESP-FPSA-2 fractional PPSA (PPSA-2/TMSA) [quantum-chemical PC]	0.098
10	ESP-PPSA-2 total charge weighted PPSA [quantum-chemical PC]	0.107
11	ESP-RNCG relative negative charge (QMNEG/QTMINUS) [quantum-chemical PC]	0.121
12	ESP-RNCS relative negative charged SA (SAMNEG*RNCG) [quantum-chemical PC]	0.439
13	ESP-RPCS relative positive charged SA (SAMPOS*RPCG) [quantum-chemical PC]	0.085
14	ESP-WPSA-1 weighted PPSA (PPSA1*TMSA/1000) [quantum-chemical PC]	0.056
15	ESP-WPSA-2 weighted PPSA (PPSA2*TMSA/1000) [quantum-chemical PC]	0.085
16	FPSA-3 fractional PPSA (PPSA-3/TMSA) [Zefirov's PC]	0.093
17	image of the Onsager−Kirkwood solvation energy	0.773
18	molecular volume	0.083
19	PPSA-3 atomic charge weighted PPSA [quantum-chemical PC]	0.096
20	principal moment of inertia A	0.056
21	RPCS relative positive charged SA (SAMPOS*RPCG) [quantum-chemical PC]	0.276
22	rotational enthalpy (300 K)/no. of atoms	0.065
23	rotational entropy (300 K)/no. of atoms	0.092
24	rotational heat capacity (300 K)/no. of atoms	0.065
25	total dipole of the molecule	0.466
26	total entropy (300 K)/no. of atoms	0.105
27	total hybridization comp of the molecular dipole	0.369
28	total point-charge comp of the molecular dipole	0.067

Table 9. Continued

no.	descriptor	significance
29	translational enthalpy (300 K)/no. of atoms	0.065
30	translational entropy (300 K)/no. of atoms	0.074
31	translational heat capacity (300 K)/no. of atoms	0.065
32	WPSA-3 weighted PPSA (PPSA3*TMSA/1000) [quantum-chemical PC]	0.345
33	XY shadow	0.136
34	zero-point vibrational energy	0.148

quantum-chemical descriptor, 7 (ESP-RPCG relative positive charge (QMPOS/QTPLUS) [quantum-chemical PC]), was not significant for herbicides because it was not able to make a difference between fungicide and herbicide. It was removed from candidate sets. The screening rules for herbicides involves seven descriptors: average bonding information content (order 1), internal enthalpy (300 K)/number of atoms, Kier and Hall index (order 2), PNSA-3 atomic charge weighted PNSA [Zefirov's PC], PPSA-2 total charge weighted PPSA [quantum-chemical PC], total heat capacity (300 K), and vibrational entropy (300 K)/number of atoms, and their data ranges are listed in **Table 23**. They were tested by three test sets, and the best result is listed in **Table 24**.

Information listed in **Table 22** shows that the accuracy of the set of screening rules for leads of herbicides was 89%. Those for leads of fungicides and insecticides were 74 and 63%, respectively. It was more significant for herbicide leads than the others because it only made 11% error for this type of lead compounds, but 26 and 37% errors for the other two types of leads, respectively.

Three test results listed in **Tables 18**, **21**, and **24**, respectively, show that the accuracies of the three sets of screening rules were >80% for their objects, respectively, but unsatisfactory performance was seen for screening out the others. The possible reasons could be proposed as follows. (1) All compounds used in this work had only one of the three confirmed and published activities. However, it is possible for them to have more than one kind of activity. (2) The analysis methods used in this work would not be perfect. (3) The descriptors calculated by

Table 10. 35 Descriptors in the Candidate Set of the Screening Rules

no.	descriptor	type
1	average bonding information content (order 1)	topological
2	average information content (order 1)	topological
3	average structural information content (order 1)	topological
4	bonding information content (order 0)	topological
5	ESP-DPSA-3 difference in CPSAs (PPSA3-PNSA3) [quantum-chemical PC]	quantum-chemical
6	ESP-PNSA-3 atomic charge weighted PNSA [quantum-chemical PC]	quantum-chemical
7	ESP-RPCG relative positive charge (QMPOS/QTPLUS) [quantum-chemical PC]	quantum-chemical
8	FNSA-2 fractional PNSA (PNSA-2/TMSA) [quantum-chemical PC]	quantum-chemical
9	FNSA-2 fractional PNSA (PNSA-2/TMSA) [Zefirov's PC]	electrostatic
10	FPSA-2 fractional PPSA (PPSA-2/TMSA) [quantum-chemical PC]	quantum-chemical
11	final heat of formation	thermodynamic
12	final heat of formation/no. of atoms	thermodynamic
13	gravitation index (all pairs)	constitutional
14	information content (order 0)	topological
15	information content (order 2)	topological
16	internal enthalpy (300 K)/no. of atoms	thermodynamic
17	internal entropy (300 K)	thermodynamic
18	internal entropy (300 K)/no. of atoms	thermodynamic
19	internal heat (300 K)	thermodynamic
20	Kier and Hall index (order 2)	topological
21	Kier and Hall index (order 3)	topological
22	PNSA-3 atomic charge weighted PNSA [Zefirov's PC]	electrostatic
23	PPSA-2 total charge weighted PPSA [quantum-chemical PC]	quantum-chemical
24	PPSA-3 atomic charge weighted PPSA [Zefirov's PC]	electrostatic
25	RPCG relative positive charge (QMPOS/QTPLUS) [Zefirov's PC]	electrostatic
26	Randic index (order 0)	topological
27	structural information content (order 0)	topological
28	total heat capacity (300 K)	thermodynamic
29	total molecular one-center E-N attraction/no. of atoms	quantum-chemical
30	vibrational enthalpy (300 K)	thermodynamic
31	vibrational enthalpy (300 K)/no. of atoms	thermodynamic
32	vibrational entropy (300 K)/no. of atoms	thermodynamic
33	WNSA-3 weighted PNSA (PNSA3*TMSA/1000) [quantum-chemical PC]	quantum-chemical
34	WPSA-2 weighted PPSA (PPSA2*TMSA/1000) [quantum-chemical PC]	quantum-chemical
35	ZX shadow	geometrical

Codessa would not be sufficient for finding perfect screening rules. Other software will be applied in our future works.

In principle, satisfactory screening rules should be sensitive for a specific type of compounds and dormant for others. Herein, we proposed three sets of screening rules, which include seven descriptors listed in **Tables 18**, **21**, and **24**, for leads of fungicides, insecticides, and herbicides, respectively. Their accuracies were 82, 83, and 89%, respectively. In principle, the three screening rules only could be used to satisfactorily estimate fungicide leads, insecticide leads, and herbicide leads, respectively. They were necessary conditions for a compound to become a fungicide lead, insecticide lead, or herbicide lead and a complement of screening rules proposed previously by us and Tice.

In any case, these rules can be used to aid researchers in virtual screening and provide references in pesticide design.

Table 11. Data Range of Candidate Screening Rules for Leads of Fungicides, Insecticides, and Herbicides

	data range		
no.[a]	fungicide	insecticide	herbicide
1	[0.5, 0.785)	[0.455, 0.725)	[0.5, 0.8)
2	[2.68, 3.82)	[2.74, 3.64)	[2.68, 4.06)
3	[0.53, 0.782)	[0.47, 0.722)	[0.494, 0.794)
4	[6.5, 18)	[9, 18.5)	[9.5, 20.5)
5	[48, 132)	[56, 140)	[60, 168)
6	[−65, −15)	[−77.5, −17.5)	[−92.5, −15)
7	[0.062, 0.271)	[0.062, 0.304)	[0.073, 0.282)
8	[−1.11, −0.2)	[−1.46, −0.2)	[−1.46, −0.13)
9	[−0.345, −0.06)	[−0.375, −0.06)	[−0.465, −0.09)
10	[0.6, 2.1)	[0.8, 2.9)	[0.6, 2.8)
11	[−155, 85)	[−245, 25)	[−215, 40)
12	[−5, 2.6)	[−7.4, 1)	[−7.4, 0.6)
13	[2200, 6200)	[2000, 6800)	[2200, 7400)
14	[39, 102)	[39, 105)	[42, 117)
15	[48, 192)	[60, 216)	[66, 210)
16	[248, 368)	[284, 422)	[272, 410)
17	[49, 141)	[81, 173)	[73, 161)
18	[2.49, 3.96)	[2.63, 4.17)	[2.56, 3.89)
19	[36, 96)	[48, 108)	[39, 105)
20	[3, 7.8)	[2.7, 9.3)	[3, 7.8)
21	[1.75, 5.25)	[1.25, 6.25)	[1.5, 5.25)
22	[−21, −2)	[−22, −3)	[−28, −4)
23	[140, 1330)	[280, 1610)	[140, 1540)
24	[4.8, 11.1)	[5.5, 12.5)	[5.5, 12.5)
25	[0.045, 0.198)	[0.09, 0.27)	[0.063, 0.216)
26	[8.5, 19)	[8.5, 20.5)	[9.5, 21)
27	[8, 18)	[9.5, 19)	[9.5, 20)
28	[40, 100)	[52, 112)	[46, 109)
29	[−190, −90)	[−180, −90)	[−210, −90)
30	[4500, 15500)	[7000, 17000)	[6000, 17500)
31	[223, 356)	[258, 391)	[258, 391)
32	[1.4, 2.65)	[1.75, 3)	[1.7, 2.75)
33	[−29, −3)	[−43, −3)	[−35, −3)
34	[0, 650)	[0, 1000)	[0, 900)
35	[27, 81)	[39, 93)	[39, 90)

[a] Corresponding to that in **Table 10**.

Table 12. 16 Groups of Descriptors Clustered by Their Definition

no.	descriptor no.[a]	definition
1	5, 6, 22, 33	relating to atomic charge
2	11, 12, 19, 28	relating to heat
3	1, 2, 3	relating to information of order 1
4	4, 14, 27	relating to information of order 0
5	16, 30, 31	relating to enthalpy
6	17, 18, 32	relating to entropy
7	7, 25	relating to relative positive charge
8	8, 9	relating to fractional partial negatively charged surface area
9	13, 35	relating to molecular shape
10	20, 21	relating to Kier and Hall index
11	23, 34	relating to weighted PPSA
12	10	fractional PPSA
13	15	information content (order 2)
14	24	atomic charge weighted PPSA
15	26	Randic index (order 0)
16	29	total molecular one-center E-N attraction/no. of atoms

[a] Corresponding to that in **Table 10**.

ABBREVIATIONS USED

Log *P*, logarithmic ratio of octanol−water partition coefficient (*P*); NHD, number of hydrogen bond donors; NHA, number of

Table 13. Most Suitable Test Result of Screening Rules for Leads of Fungicides among 41472 Combinations

screening rule[a]	fungicide[b]	insecticide[b]	herbicide[b]
1, 4, 9, 10, 12, 13, 15, 16, 18, 20, 24, 25, 26, 29, 33, 34	80%	54%	74%

[a] Screening rule represented by the number of a descriptor corresponding to that in **Table 10**. [b] If a compound has five descriptors in this screening rule that are not in the data ranges, the compound will not be a lead of fungicide.

Table 14. Most Suitable Test Result of Screening Rules for Leads of Insecticides among 41472 Combinations

screening rule[a]	fungicide[b]	insecticide[b]	herbicide[b]
2, 4, 5, 7, 8, 10, 11, 15, 17, 20, 23, 24, 26, 29, 30, 35	70%	80%	76%

[a] Screening rule represented by the number of a descriptor corresponding to that in **Table 10**. [b] If a compound has five descriptors in this screening rule that are not in the data ranges, the compound will not be a lead of insecticide.

Table 15. Most Suitable Test Result of Screening Rules for Leads of Herbicides among 41472 Combinations

screening rule[a]	fungicide[b]	insecticide[b]	herbicide[b]
1, 7, 9, 10, 13, 14, 15, 16, 20, 22, 23, 24, 26, 28, 29, 32	68%	72%	91%

[a] Screening rule represented by the number of a descriptor corresponding to that in **Table 10**. [b] If a compound has five descriptors in this screening rule which are not in the data ranges, the compound will not be a lead of herbicide.

Table 16. Test Results of the Two Electrostatic Descriptors, 24 and 25

descriptor no.[a]	fungicide	insecticide	herbicide
24	76%	82%	72%
25	72%	66%	87%

[a] Corresponding to that in **Table 10**.

Table 17. Screening Rules for Leads of Fungicides

descriptor no.[a]	data range	descriptor no.[a]	data range
1	[0.5, 0.785)	26	[8.5, 19)
9	[−0.345, −0.06)	29	[−190, −90)
12	[−5, 2.6)	33	[−29, −3)
16	[248, 368)		

[a] Corresponding to that in **Table 10**.

Table 18. Test Result of Screening Rules for Leads of Fungicides

screening rule[a]	fungicide[b]	insecticide[b]	herbicide[b]
1, 9, 12, 16, 26, 29, 33	82%	56%	70%

[a] Screening rule represented by the number of a descriptor corresponding to that in **Table 10**. [b] If a compound has two descriptors in this screening rule that are not in the data ranges, the compound will not be a lead of fungicide.

Table 19. Test Results of the Two Descriptors, 2 and 7

descriptor no.[a]	fungicide	insecticide	herbicide
2	60%	62%	62%
7	82%	77%	82%

[a] Corresponding to that in **Table 10**.

Table 20. Screening Rules for Leads of Insecticide

no. of descriptor[a]	data range	no. of descriptor[a]	data range
5	[56, 140)	24	[5.5, 12.5)
8	[−1.46, −0.2)	29	[−180, −90)
10	[0.8, 2.9)	35	[39, 93)
11	[−245, 25)		

[a] Corresponding to that in **Table 10**.

Table 21. Test Result of Screening Rules for Leads of Insecticide

screening rule[a]	fungicide[b]	insecticide[b]	herbicide[b]
5, 8, 10, 11, 24, 29, 35	60%	83%	74%

[a] Screening rule represented by the number of a descriptor corresponding to that in **Table 10**. [b] If a compound has two descriptors in this screening rule which are not in the data ranges, the compound will not be a lead of insecticide.

Table 22. Test Result of Descriptor 7

descriptor no.[a]	fungicide	insecticide	herbicide
7	80%	63%	80%

[a] Corresponding to that in **Table 10**.

Table 23. Screening Rules for Leads of Herbicides

descriptor no.[a]	data range	descriptor no.[a]	data range
1	[0.5, 0.8)	23	[140, 1540)
16	[272, 410)	28	[46, 109)
20	[3, 7.8)	32	[1.7, 2.75)
22	[−28, −4)		

[a] Corresponding to that in **Table 10**.

Table 24. Test Result of Screening Rules for Leads of Herbicides

screening rule[a]	fungicide[b]	insecticide[b]	herbicide[b]
1, 16, 20, 22, 23, 28, 32	74%	63%	89%

[a] Screening rule represented by the number of a descriptor corresponding to that in **Table 10**. [b] If a compound has two descriptors in this screening rule that are not in the data ranges, the compound will not be a lead of insecticide.

hydrogen bond acceptors; MW, molecular weight; PSA, polar surface area; K−S test, Kolmogorov−Smirnov test; ANOVA, analysis of variance; K−W test, Kruskal−Wallis test; PNSA, partial negatively charged surface area; PPSA, partial positively charged surface area.

Supporting Information Available: Pearson correlation coefficients amont 35 descriptors. This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Tice, C. M. Selecting the right compounds for screening: does Lipinski's rule of 5 for pharmaceuticals apply to agrochemicals? *Pest Manag. Sci.* **2001**, *57*, 3–16.
(2) Tice, C. M. Selecting the right compounds for screening: use of surface-area parameters. *Pest Manag. Sci.* **2002**, *58*, 219–233.
(3) Liu, B.; Yu, F.; Yao, J. H.; Liao, Q.; Fan, B. T. Screening rules of lead compounds of herbicide, fungicide and insecticide. *Chin. J. Pest. Sci.* **2007**, *9*, 220–228.
(4) Semichem, Inc. http://www.semichem.com/codessa/default.php.
(5) CODESSA Reference Manual.
(6) http://en.wikipedia.org/wiki/K-S_test.
(7) Fisher, R. A. In *Statistical Methods for Research Workers*; Oliver and Boyd: Edinburgh, Scotland, 1925.
(8) http://en.wikipedia.org/wiki/Analysis_of_variance.
(9) http://en.wikipedia.org/wiki/Wallis_statistic.
(10) Kruskal, W. H.; Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Am. Statis. Assoc.* **1952**, *47*, 583–621.

(11) http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient.
(12) Cann, A. J. In *Maths from Scratch for Biologists*; Wiley: West Sussex, U.K., 2003; pp 135−146.
(13) Tomlin, C. D. S. In *The e-Pesticide Manual*, 12th ed.; British Crop Protection Council: Surrey, U.K., 2000.
(14) In *Pesticides, An International Guide to 1800 Pest Control Chemicals*, 2md ed.; Milne, G. W. A., Ed.; Ashgate Publishing: Surrey, U.K., 2004.
(15) Patent PCT/CN03/00343.
(16) Okazawa, A.; Akamatsu, M.; Nishiwaki, H.; Nakagawa, Y.; Miyagawa, H.; Nishimura, K.; Ueno, T. Three-dimensional quantitative structure−activity relationship analysis of acyclic and cyclic chloronicotinyl insecticides. *Pest Manage. Sci.* **2000**, *56*, 509–515.
(17) http://www.epa.gov/.
(18) http://www.cas.org/SCIFINDER/SCHOLAR.
(19) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
(20) Kier, L. B. Use of molecular negentropy to encode structure governing biological activity. *J. Pharm. Sci.* **1980**, *69*, 807–810.
(21) Stankevich, I. V.; Stankevich, M. I.; Zefirov, N. S. Topological indices in organic chemistry. *Russ. Chem. Rev.* **1988**, *57*, 191–208.
(22) El-Basil, S.; Randic, M. Equivalence of mathematical objects of interest in chemistry and physics. *Adv. Quantum Chem.* **1992**, *24*, 239–290.
(23) Rohrbaugh, R. H.; Jurs, P. C. Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. *Anal. Chim. Acta* **1987**, *199*, 99–109.
(24) Karelson, M. In *Molecular Descriptors in QSAR/QSPR*; Wiley: New York, 2000.
(25) Osmialowski, K.; Halkiewicz, J.; Kaliszan, R. Quantum chemical parameters in correlation analysis of gas−liquid chromatographic retention indices of amines. *J. Chromatogr., A* **1986**, *361*, 63–69.
(26) Stanton, D. T.; Jurs, P. C. Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure−property relationship studies. *Anal. Chem.* **1990**, *62*, 2323–2329.
(27) Bodor, N.; Gabanyi, Z.; Wong, C.-K. A new method for the estimation of partition coefficient. *J. Am. Chem. Soc.* **1989**, *111*, 3783–3786.
(28) Sannigrahi, A. B. AB initio molecular orbital calculations of bond index and valency. *Adv. Quantum Chem.* **1992**, *23*, 301–351.
(29) SPSS Inc. http://www.spss.com/.
(30) Lead Compound Screening System (CISOC-LSS), Registration Number (China): 096088.

Received for review July 29, 2009. Revised manuscript received November 15, 2009. Accepted November 19, 2009. This work received financial support from the Ministry of Science and Technology of China (Grants 2003CB114400 and 2006AA02Z339), the National Natural Science Foundation of China (Grant 20572081), and the Ministry of Environmental Protection of China (Grant 200709046).