Structure-Activity Relationships on Pesticides: A Development in Methodology and Its Software System

Jiaju Zhou,* Qian Xie, Dunming Sun, Guirong Xie, Lingxiao Cao, and Zhihong Xu

Laboratory of Computer Chemistry (LCC), Institute of Chemical Metallurgy, Chinese Academy of Sciences, P.O. Box 353, Beijing 100080, China

Received July 28, 1992

An overview of the project "Computer-Aided Molecular Design of Pesticide" running in LCC, Chinese Academy of Sciences, is presented. In this paper, the authors propose a structural frame model (SFM) method for numeralization of structural factors and report a user-friendly software named "Computer-Aided Screening of Agricultural Chemicals" (CASAC) developed by LCC, including a prototype of structure-activity database (S-ADB) to support SAR/QSAR studies and computer-aided molecular design. Some typical case studies on herbicides, insecticides, and chemical hybridizing agents have been represented. The paper is focused on the SFM method for superimposing structural information and noteworthy functions of the CASAC system (version 1991).

INTRODUCTION

In 1989, a project "Computer-Aided Molecular Design of Pesticide" was initiated in LCC. The main task was to develop a user-friendly software system to meet the urgent needs coming from new pesticide researchers, who have found that discovering a new pesticide becomes more and more resource- and time-consuming.

Depending on whether the structure of biomacromolecules, which interact with the small molecules of pesticides, is known or not, there are two approaches to investigate the structure-activity relationships (SAR). If the structure of biomacromolecules is known, we may find a drug more rationally. It is the first approach, but in many cases, the structure of biomacromolecules is hard to know. In the field of pesticide science, unfortunately, at most conditions we do not know the structure. So we have to develop another approach to approach new pesticides from known chemicals. As a software of computer-aided screening, the CASAC system starts from the activity of known compounds or commercial pesticide molecules.

In the research of QSAR, the Hansch-Fujita approach[1] is the most remarkable and popular one. The core of the Hansch-Fujita approach is that they proposed three kinds of key physicochemical parameters—the hydrophobic parameters of molecules, the electronic parameters, and the stereoeffect parameters of substituents—which associate tightly with the bioactivity of a drug molecule.

Many successes have been achieved with the Hansch method in last 3 decades. However, a user-friendly software of computer-aided molecular design (CAMD) is still needed. The reasons are as follows:

1. During the screening and biotesting processes, most of the activity data produced are qualitative data, which can not fit in Hansch-Fujita method. The Hansch-Fujita method is a quantitative method; it requires quantitative activity data, such as, IC_{50}, ED_{50}.
2. The selection of important physicochemical parameters is experience-based or even art-based.
3. In some cases, it is quite difficult to get the parameters of unusual substituents or molecules.
4. For many systems investigated, it has been found that the important parts are not only substituents but also some parts of the molecule. In such cases, the physicochemical parameters of substituents alone are not enough for SAR study.
5. In traditional QSAR, people mainly use the regression method to construct QSAR equations. We think that lots of other mathematical methods also have great potential in SAR/QSAR researches.

With the reasons mentioned above, we attempted to develop a generalized method and a user-friendly software for structural information representation and extraction to help pesticide/drug screenings and activity predictions.

STRUCTURAL FRAME MODEL (SFM) METHOD

SFM is a utility for transforming structural information into numerical information. The main idea about SFM is, for a set of compounds including both training set and unknown set of compounds, to superimpose the common backbone of all compounds together at first and, then, consider all changeable structural factors and design a structural frame composed of the common backbone and frame sites. Here, structural factor is an expanding concept of substituent and can be any part of a molecule. Using this method, one can easily realize numeralization of structural information.

1. Method. The working procedure of SFM is

(a) define and draw a SFM
(b) define a set of "frame sites"
(c) write a code sequence about structural factors for each compound according to its structure
(d) put the code sequence and bioactivity data together and form an input file for the CASAC system

2. Illustration. As an example, the 126 sulfonyl urea compounds are taken from a Du Pont Patent—U.S. 4854962.[2] In the set of compounds, the common backbone is the sulfonyl urea bridge in the middle, a variety of substitutive benzoheterocycles are attached to the sulfonyl group, and substitutive pyrimidine or triazine rings are connected to the other end of the sulfonyl urea bridge. Figures 1 and 2 show the structural formulas of the variable parts of the compounds.

(a) SFM of the Sulfonylurea Compounds. According to the structural variance, the SFM is defined as Figure 3. The SFM resembles the traditional structural formula and could be considered as a "superimposed structural formula" con-

0095-2338/93/1633-0310$04.00/0

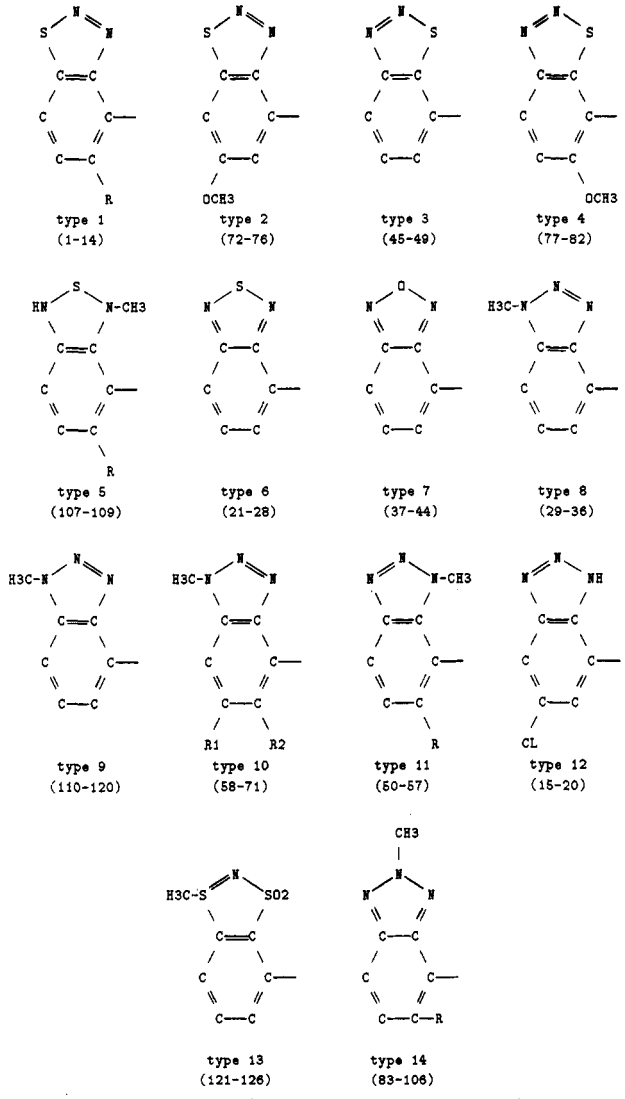


Figure 1. Fourteen types of benzoheterocycles on the left side of 126 sulfonyl urea compounds. The compound numbers are shown in parentheses.

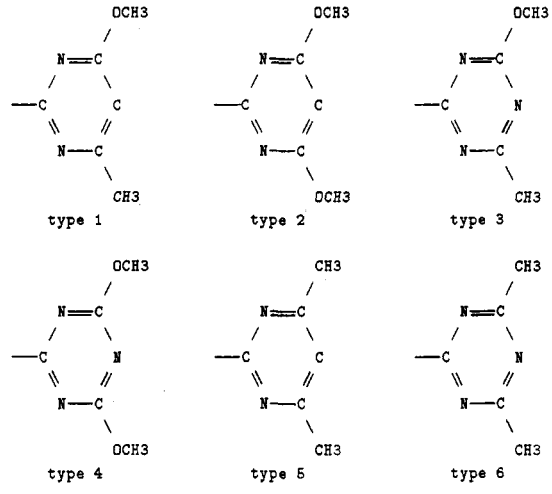


Figure 2. Six main types of substituted heterocycles on the right side of 126 sulfonylurea compounds. In the 126 compounds, compounds 110–120 have different substituents on frame site S2 and S4, see Table II.

sisting of frame sites and the backbone. A frame site can be replaced with an atom, a functional group, or any part of the molecule. We call them "structural factors". All of the structural factors at frame sites L0, L1, M1, M12, and others have definite chemical meanings.

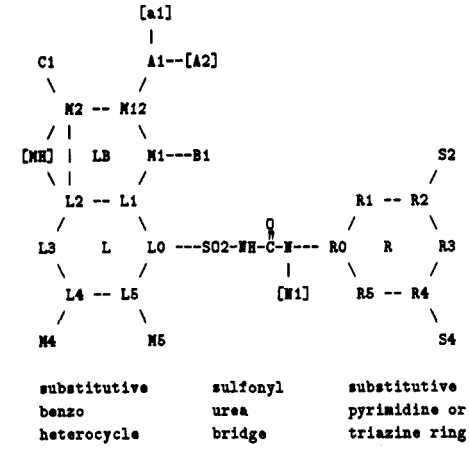


Figure 3. SFM of the sulfonyl urea compounds. Frame sites [MH], [N1], [A2], and [a1] are not discussed in this section but are used in the section of case studies. See note in ref 7.

(b) Definition of Frame Sites. A frame site is a generalized structural factor. In Figure 3, we define R and L as the ring on the right and left side, respectively. The R ring contains six frame sites R0, R1, R2, R3, R4, and R5. The L ring also consists of six frame sites L0, L1, L2, L3, L4, and L5. The LB means the five-membered ring consisting of L1, L2, M1, M2, and M12. A1, B1, C1, M4, M5, S2, and S4 are associated with frame sites M12, M1, M2, L4, L5, R2, and R4, respectively. M12 determines whether a benzo five-membered ring is present.

In these frame sites, considering the current 126 compounds, there are 11 unchanged sites—L0, L1, L2, L3, L4, L5, R0, R1, R2, R4, and R5—and we fix them reasonably from now on. The remaining 11 sites are numbered as follows (the order is not important):

1	2	3	4	5	6	7	8	9	10	11
S2	S4	R3	M1	M2	M12	M4	M5	A1	B1	C1

(c) Code Sequence. For numeralization of structural factors, we have established a parameter database PARABAS which contains over 300 structural factors now. In PARABAS, each structural factor is expressed with a four-character code. For each compound within the SFM, we replace each changeable frame site with the code of the corresponding structural factor, which forms a code sequence. For example, the first five compounds have the code sequences shown in Chart I.

Here, CH3, OCH3, CL, ~CH~, ~N~, =N–, –N=, –S–, and – are structural factor codes. CH3 is methyl, OCH3 represents methoxy, ~CH~ and ~N~ imply the CH group and the nitrogen atom on an aromatic ring, respectively, and so forth. The code "–" means null. The code as intuitional as possible.

(d) The Only Input File of CASAC System. In order to form the input file of the CASAC system, the operation needed is simply to insert one or more columns of bioactivity data in front of the code sequences. In biotesting, the activity data of the first five compounds are 7, 9, A, 7, and 0, respectively. So the input file for them are shown in Chart II.

Remarks: target plant is Morning glory (*Ipomoea spp.*), rate = 0.05 Kg/ha, and the schedule of treatment is postemergence.

It should be indicated that the CASAC system needs only this file from the user. Other parameters are produced by the system.

Those rules about SFM definition of structural factors have wide applications. With them, one can process construction

Chart I

No.	1	2	3	4	5	6	7	8	9	10	11
	S2	S4	R3	M1	M2	M12	M4	M5	A1	B1	C1
001	CH3	CH3	~CH~	=N—	—S—	—N=	H	OCH3	–	–	–
002	OCH3	CH3	~CH~	=N—	—S—	—N=	H	OCH3	–	–	–
003	OCH3	OCH3	~CH~	=N—	—S—	—N=	H	OCH3	–	–	–
004	OCH3	CL	~CH~	=N—	—S—	—N=	H	OCH3	–	–	–
005	CH3	CH3	~N~	=N—	—S—	—N=	H	OCH3	–	–	–

Chart II

001	7	CH3	CH3	~CH~	=N—	—S—	—N=	H	OCH3	–	–	–
002	9	OCH3	CH3	~CH~	=N—	—S—	—N=	H	OCH3	–	–	–
003	A	OCH3	OCH3	~CH~	=N—	—S—	—N=	H	OCH3	–	–	–
004	7	OCH3	CL	~CH~	=N—	—S—	—N=	H	OCH3	–	–	–
005	0	CH3	CH3	~N~	=N—	—S—	—N=	H	OCH3	–	–	–

of ordinary small pesticide molecules with branches and rings. For instance, the six structure factors, L0, L1, L2, L3, L4, and L5, actually define all rings from three-membered to six-membered ones, including saturated, unsaturated, aromatic rings, and a variety of heterocycles. In this example, if L3 = "–" (null), a five-membered ring is defined; furthermore, when L3 = L4 = "–", the L0, L1, L2, and L5 form a four-membered ring. The structural factor MH obeys a similar rule and has the same function. In this way, six variables can express many structural changes.

It is worthy to mention that after the job is done the CASAC will do many jobs for users, even the analyses of the output results.

CASAC: COMPUTER-AIDED SCREENING OF AGRICULTURAL CHEMICALS SYSTEM

The CASAC (version 1991) is a user-friendly software system. It is mainly comprised of four subsystems: (1) structural information processors—the structural factor codes and parameter database (PARABAS) and related programs; (2) correlative tools—the multivariable statistics package (statistical tools) and the artificial neural network (ANN tools); (3) resources of activity information—the pesticide structures–activities database (S–ADB); and (4) computational chemistry (QC and MM) package. Figure 4 shows the conceptual framework of the CASAC software system.

Using the CASAC system, one can easily do SAR/QSAR studies, and CAMD works with higher efficiency and lower resource consumption. The main techniques adopted in CASAC are automatic numeralization of all structural code sequences, automatic detection, and determination of important structural factors. We have developed several methods to numeralize structural factors. In other words, it is not necessary to manually select important structural factors, prepare parameters by experimental values from handbooks or references, or calculate them using complicated methods, like the CLOGP for hydrophobic parameters. Certainly, the users can have their own parameters used in CASAC system, and we have developed such an interface.

1. Structural Information Processors. Because we do not know in advance which part of a compound may be important

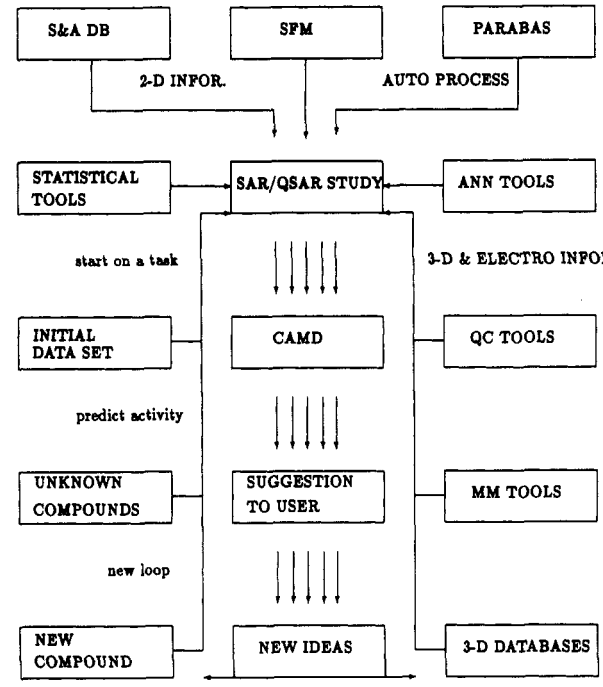


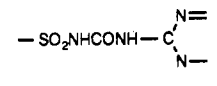
Figure 4. Conceptual framework of the CASAC software system. In all cases, the start point of a task must be from a set of experimental bioactivity data.

for its activity, a default path for CASAC has been designed, in which three kinds of parameters, OI, EN, and MR, are used for all structural factors. With the help of the PARABAS and related programs, OI, EN, and MR values of all structural factors are used as independent variables (descriptors) and put into a file for the following analyses and/or ANN calculation.

OI is the ratio of the organic property and the inorganic property of a compound. EN and MR are group electronegativity and group refractivity of a structural factor, respectively.

The first parameter OI is a substitute of log P. The concept of organic property value and inorganic property value were given by a Japanese chemist many years ago.[3] We have studied the correlativity between OI and log P within many sets of compounds, each set of compounds can fit in one frame. For

Table I. Examples of Structural Factors Used in CASAC System

structural factor	structure	code in PARABAS
C atom with 4 single bonds	>C<	>C<
CH group with 3 single bonds	–CH<	–CH<
methyl	–CH3	CH3
methoxy	–OCH3	OCH3
ethoxy	–OCH2CH3	OET
ethylthio	–SCH2CH3	SET
carbonyl	–CO–	T
sulfonyl(=sulfuryl)	–SO2–	SO2
carboxy	–COOH	COOH
tert-butyl	–C(CH3)3	T-BU
bromo	–Br	BR
triphenylmethyl	–C(C6H5)3	TAAA
CH group in aromatic ring		~CH~
C atom in aromatic ring		~C~–
N atom in aromatic ring		~N~
ureido	–NHCONH2	NTN
ureylene	–NHCONH–	NTN–
phenyl	–C6H5	C7H5
cyclohexyl	–C6H11	CYHE
a bridge of sulfonylurea with a part of pyrimidine or triazine ring		no
null	/	"–"

the 10 systems tested, the structures have different backbones, and values of the correlativity criteria R are from 0.891 to 0.978. These results prove that the correlativity between OI and log P is good enough. For details, please see ref 4. On the basis of this study, we come to a decision, using OI to replace the hydrophobic parameter log P in CASAC.

The second derived parameter is EN. One of the authors of this paper, Q.X., improved and expanded the EN calculating method suggested by Han[5] (for simple substituents only) and proposed a new method[6] for any kind of structural factors. In Xie's method, the self-consistence idea has been successfully used.

PARABAS contains numerical data about OI, EN, and MR parameters of over 300 structural factors. The system transfers code sequences of a compound in an input file into real number vectors. For unusual groups, the user is asked to input a connect table of this special group. After that, the system will calculate the parameter values of the new group. The advantage of the full-automation of getting the values of parameters is clear.

In the project, we emphasize the concept of the "structural factor", which is different from that of the substituent or group. A structural factor may be an atom, a substituent, or, generally speaking, any part of a molecule. For example, in cases of one carbon with 0–3 hydrogen atoms: all of >C<, >CH–, $-CH_2-$, $-CH_3$, >C=, –CH=, $CH_2=$, =C=, #C–, #CH, ~CH~, and ~C~– may be a structural factor. In the CASAC system, the structural factor is a flexible and vivid concept. Because any part, one or more, of a molecule may be an activity part. We hope that by using some powerful mathematical methods, such as factor analysis (FA), these activity parts will be found out at least in the sense of statistical correlation or modeling. Some simple examples of structural factors are shown in Table I.

2. Correlative Tools. By functions, the CASAC (version 1991) contains three groups of multivariable statistics as well as artificial intelligence tools.

In the first group, principal components analysis (PCA), factor analysis (FA), and hierarchical clustering analysis (HCA) are nonsupervisor methods and do not need bioactivity data, which can analyze the feature of descriptors and eliminate some redundant descriptors. Of course, these methods can also be used for classification when bioactivity data are available.

Stepwise discriminant analysis (SDA), discriminant analysis with constellation graph (DACG), nonlinear mapping (NLM), HCA, and artificial neural network (ANN) belong to the second group. They are used for the qualitative activity data.

The last group involves the traditional Hansch method, multivariable regression, and other QSAR methods. It can only treat quantitative activity data. Novel ANN methods can treat both qualitative and quantitative activity data.

All programs are written in FORTRAN language and can run on both VAX-11/780, IBM/PC 286, and compatible computers. Some subroutines are taken from the SASD package (a Package for Statistical Analysis of Stochastic Data[7]). Meanwhile a FORTRAN plot program is developed especially for pattern recognition to display NLM, DACG, and PCA patterns on the screen or plot them on the LVP16 plotter.

For all correlative software tools, details of the methods, and applications to pesticide molecular design, please read Dunming Sun's paper[8] (or other proper publications).

From a user's point of view some highlights of the CASAC system are briefly mentioned here. In the experience of the authors, the SDA program is very powerful for classification and prediction of activities, and NLM or DACG plots on the screen are quite efficient and intuitive also. The PCA is usually done as a pre-step of a job, and the FA has notable potential to find some undiscovered activity structural factors in a frame. The ANN approach is suitable to process the system with high nonlinear distribution but often takes longer CPU time than most of other methods.

3. Structure–Activity Database (S–ADB). Generally speaking, the starting point in SAR and CAMD work is a set of experimental activity data either from the laboratories or the literature. At present, the screening method is still the main method to find new pesticides. In the literature, especially in patents, there are a mass of data involving various bioactivities of all sorts of compounds. Those are very precious resources for SAR study. As a part of the CASAC system, we have developed a prototype of structure–activity database—S–ADB. The most important feature of S–ADB is that it contains both multitarget bioactivity data and structural information. This feature makes S–ADB become a convenient data source for structure–activity relationship studies. S–ADB is a developing database in LCC. Now there are about 3000 herbicide compounds, 1000 fungicide compounds, and 1000 of the newest bioactivity compounds in S–ADB. Most of original data are from patents. After selection, evaluation, recompilation of activity data, and writing of topological connected tables for each compound, LCC developed the prototype of a new type of database.

Table II lists partly a data file of S–ADB with the typical format of the CASAC system. The left part in each row is activity data, while the right part is the structural information in code sequences. Each row corresponds to a compound and each column within the activity data region (left part) contains all the compounds' activity ranks to a particular target plant. For instance, column 1 corresponds to Morning glory (*Ipomoea spp.*) codes by MORN (read the code in the heading of the table vertically), column 2 to Cocklebur (*Xanthium pensylvanicum*) coded by COCK, and so on. From both the style and the content of the file, one will concretely understand that biological and chemical information in S–ADB are all

Table II. Bioactivities and Structural Factors of 205 Herbicidal Sulfonylurea Analogs[a]

	Bioactivities(i) (i= 1 to 50) *		Structural Factors(j)** (j= 1 to 15)														
	<--- POSTemergence -----><---- PREemergence ----->																
	12345678901234567890123456789012345678901234567890																
	***																
No	MCVNCBWWCSRSCSCSBBCGDNNNNMCVNCBWWCSRSCSCSBBCGDNNNN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	OOEURAIHOOIOHUOIAUAIOUUUUOOEURAIHOOIOHUOIAUAIOUUUU																
	RCLTARLERYCREGTCRSSAWSSSSRCLTARLERYCREGTCRSSAWSSSS	S2	S4	R3	M1	M2	M12	M4	M5	A1	B1	C1	A2	a1	MH	M1	
	NKVSBNDANBEGAATKLHSNNEEEENKVSBNDANBEGAATKLHSNNEEEE																
001	779999999989999A/////////3/5A67899599947/////////	CH3	CH3	"CH"-N=	-S-	-N=	H		OCH3-		-	-	-	-	-	-	
002	9A999A999A9999A/////////657A99899799979/////////	OCH3CH3		"CH"-N=	-S-	-N=	H		OCH3-		-	-	-	-	-	-	
003	AA999999A99999A/////////958989899799969/////////	OCH3OCH3		"CH"-N=	-S-	-N=	H		OCH3-		-	-	-	-	-	-	
004	7A9999899999999/////////94845804959A958/////////	OCH3CL		"CH"-N=	-S-	-N=	H		OCH3-		-	-	-	-	-	-	
005	000009999589802/////////0000000730990O2/////////	CH3	CH3	"N" -N=	-S-	-N=	H		OCH3-		-	-	-	-	-	-	
006	7/6099999999986/////////9878898997A9989/////////	OCH3CH3		"N" -N=	-S-	-N=	H		OCH3-		-	-	-	-	-	-	
007	595899999999999A/////////810A849992AA897/////////	OCH3OCH3		"N" -N=	-S-	-N=	H		OCH3-		-	-	-	-	-	-	
008	AAAA992AA9999AA/9//A////979A894999AA999/2//9/////	CH3	CH3	"CH"-N=	-S-	-N=	H		H	-	-	-	-	-	-	-	
009	AAAAAA99A999AAA/9//A////7995897999A9999/9//9/////	OCH3CH3		"CH"-N=	-S-	-N=	H		H	-	-	-	-	-	-	-	
010	AAAA9999A99AAAA/9//A////9899999699999999/8//9/////	OCH3OCH3		"CH"-N=	-S-	-N=	H		H	-	-	-	-	-	-	-	
011	A9A99A59A9999AA/4//A////9/8679389899798/7//8/////	OCH3CL		"CH"-N=	-S-	-N=	H		H	-	-	-	-	-	-	-	
012	A9A5A909A9999AA/9//A////99999949999AA999/9//9/////	OCH3CH3		"N" -N=	-S-	-N=	H		H	-	-	-	-	-	-	-	
013	AAA99929A9996AA/9//A////979A9929989A879/9//9/////	OCH3OCH3		"N" -N=	-S-	-N=	H		H	-	-	-	-	-	-	-	
014	9AA499899999996A/7//9////9/87783397A9907/0//8/////	OET	NHM	"N" -N=	-S-	-N=	H		H	-	-	-	-	-	-	-	
015	000000000000000/0//0////000000000000000/0//0/////	CH3	CH3	"CH"-N<	-N=	-N=	CL		H	-	H	-	-	-	-	-	

[a] Compounds 001–126 are from U.S. Patent 4854962 (1989), and compounds 201–280 are from U.S. Patent 4741761 (1986). Notes: *, number of system; **, number of frame sites; ***, code of plants (first four characters of the target plant names and these characters must be read by column, for example, the MORN means Morningglory (*Ipomoea* spp.), and especially NUSE implies that the column is not used at present. Special codes of structural factors: NMCN, NHCH2CN; OE#C, OCH2C#CH; COM2, CH(OCH3)2; NMMQ, N(CH3)CH2CN; NHM, NHCH3; OEF3, OCH2CF3; OCF2, OCF2H; CL2, <ClCl for same C; MOM, CH2OCH3; NM2, N(CH3)2; and OCMD, OC(CH3)=CH- to next C as a ring.

Table III. Distribution of Activity Classes of 205 Compounds for 13 Target Plants and Two Modes of Usage[a]

no. of system	mode of usage	target of plants	profile of activities by compounds' numbers											
			/	0	1	2	3	4	5	6	7	8	9	A
1	POST	MORNingglory	0	54	7	14	7	4	10	2	12	19	47	29
2	POST	COCKlebur	2	37	5	5	13	9	10	4	12	9	52	47
4	POST	NUTSedge	2	68	0	6	3	5	9	2	6	26	60	18
5	POST	CRABgrass	0	71	2	11	13	6	10	8	7	12	59	6
6	POST	BARNyardgrass	0	48	1	10	5	2	3	2	10	12	101	11
7	POST	WILD-oats	0	67	2	9	6	5	5	2	4	15	90	0
8	POST	WHEAt	0	67	1	9	1	4	7	3	3	6	91	13
9	POST	CORN	0	41	0	5	3	3	6	4	6	10	89	38
10	POST	SOYBean	0	37	3	9	4	2	9	4	15	24	97	1
11	POST	RICE	0	31	0	10	3	4	4	1	4	12	135	1
12	POST	SORGhum	0	36	0	4	3	2	11	6	9	8	110	16
14	POST	SUGAr-beets	1	39	3	9	7	1	8	7	4	12	83	31
15	POST	COTTon	5	39	4	10	5	3	8	7	15	14	67	28
26	PRE	MORNingglory	0	87	5	14	8	4	7	6	5	11	54	4
27	PRE	COCKlebur	15	75	4	18	9	5	6	2	6	16	49	0
29	PRE	NUTSedge	0	105	0	4	3	4	6	2	8	13	18	42
30	PRE	CRABgrass	6	95	2	12	11	4	11	8	10	12	34	0
31	PRE	BARNyardgrass	0	89	2	9	5	6	6	2	9	12	63	2
32	PRE	WILD-Oats	0	93	1	12	7	4	7	4	8	19	48	2
33	PRE	WHEAt	0	95	0	6	4	4	5	1	4	15	48	23
34	PRE	CORN	0	75	2	9	5	6	1	1	7	13	76	10
35	PRE	SOYBean	0	78	4	24	13	10	8	11	15	10	32	0
36	PRE	RICE	0	65	0	6	5	1	10	2	12	6	34	64
37	PRE	SORGhum	0	66	0	7	1	2	8	4	4	10	75	28
39	PRE	SUGAr-Beets	1	64	3	6	9	10	8	2	16	18	54	14
40	PRE	COTTon	8	90	3	13	7	4	6	3	10	20	40	1

[a] All the numbers in the table's body are the compound numbers to every activity class. "/" means the number of compounds without biotesting.

computerized and very condensed. Furthermore, it is the only input file that the CASAC system needs from users.

Even though S–ADB is only a prototype, it has became a very useful tool for both software system designers and users to do SAR study, structure optimization, new drug design, and software development. As the software designers, we can conveniently extract good data sets in which the activity data have a well-distributed profile and the structural variation is well-dispersed. They are good for some novel calculating programs to test and verify their efficiency. For users of

Table IV. Summary of Classification for 26 Systems[a]

no.	plant	mode of use	N_e/N_t	error %	D2
1	Morning glory	POST	14/130	10.8	7.86
2	Cocklebur	POST	11/136	8.1	5.61
4	Nutsedge	POST	7/146	4.8	9.10
5	Crabgrass	POST	13/136	9.6	10.09
6	Barnyard grass	POST	7/160	4.4	13.30
7	Wild-Oats	POST	20/157	12.7	6.52
8	Wheat	POST	14/167	8.4	6.45
9	Corn	POST	11/168	6.5	8.93
10	Soybean	POST	7/135	5.2	7.15
11	Rice	POST	6/167	3.6	18.22
12	Sorghum	POST	8/162	4.9	14.27
14	Sugar-Beets	POST	6/153	3.9	8.82
15	Cotton	POST	9/134	6.7	8.32
26	Morning glory	PRE	13/145	9.0	9.11
27	Cocklebur	PRE	12/124	9.7	9.45
29	Nutsedge	PRE	28/165	17.0	3.33
30	Crabgrass	PRE	9/129	7.0	10.14
31	Barnyardgrass	PRE	17/154	11.0	7.63
32	Wild-Oats	PRE	7/143	4.9	19.60
33	Wheat	PRE	18/166	10.8	7.35
34	Corn	PRE	12/161	7.5	10.06
35	Soybean	PRE	11/110	10.0	6.57
36	Rice	PRE	13/163	8.0	9.39
37	Sorghum	PRE	14/169	8.3	8.16
39	Sugar-Beats	PRE	14/132	10.6	6.76
40	Cotton	PRE	8/31	6.1	10.39
	average			8.1	

[a] N_e is number of misclassified compounds. N_t is the total number of compounds in the system. D2 is Mahalanobis distance between two classes. In SDA method, the criterions of variable selection (F1) and removement (F2) are 1.5.

Table V. Summary of Activity Predictions for 13 Postemergence Systems[a]

no.	plant	$N1_t$	$N1_e$	$N2_t$	$N2_e$	N_t	N_e	agreed %
1	Morning glory	76	12	54	12	130	24	81.5
2	Cocklebur	98	16	37	14	135	30	77.8
4	Nutsedge	78	9	68	13	146	22	85.0
5	Crabgrass	65	6	71	10	136	16	88.2
6	Barnyard grass	111	14	48	6	159	20	87.4
7	Wild-Oats	90	10	67	16	157	26	83.4
8	Wheat	104	5	67	15	171	20	88.3
9	Corn	127	4	41	12	168	16	90.5
10	Soybean	97	7	37	11	134	18	86.6
11	Rice	136	3	31	6	167	9	94.6
12	Sorghum	126	5	36	8	162	13	92.0
14	Sugar-Beets	114	5	39	10	153	15	90.2
15	Cotton	95	13	39	9	143	22	83.6
	average							86.9

[a] $N1_t$ is the total number of compounds in class 1. $N1_e$ is the number of misclassified compounds in class 1. $N2_t$ is the total number of compounds in class 2. $N2_e$ is the number of misclassified compounds in class 2. N_t is the total number of compounds. N_e is the number of misclassified compounds. In SDA, both F1 and F2 are 1.5.

CASAC, we believe that, if they combine their own experimental data with some data of the S–ADB to form an enlarged data set, they will get great benefits. With the help of the enlarged data set and the CASAC system, many unnecessary syntheses will be avoided.

4. Computational Chemistry (QC and MM) Package. This package is absolutely necessary for any three-dimensional SAR/QSAR study and three-dimensional structural databases. It contains Gaussian90, AMPAC/MOPAC, MM2, CSD, ICSD, etc.

CASE STUDIES

In this section, we will show some examples of CASAC's applications on different kinds of agricultural chemicals.

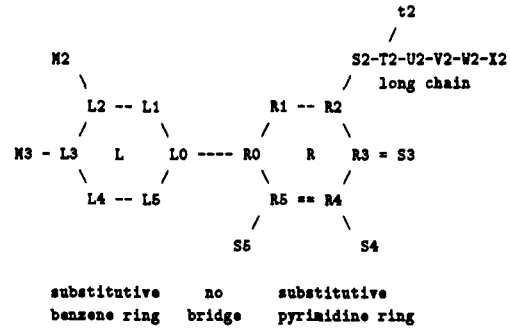


Figure 5. SFM of the chemical hybridizing agent (CHA) analogs. Frame sites L0, L1, L2, L3, L4, L5, R0, R1, R2, R3, R4, and R5 are the same for all 40 compounds. R0 is –N<, and R1 is –N=. S2 is fixed as a carbonyl. There are 11 variable sites, S4, S5, S3, M3, M2, T2, U2, V2, W2, X2, and t2. In the set of compounds, frame sites are replaced by structural factors: S4 = H, CL, Br, C3H7; S5 = CH3, C2H5, C4H9; S3 = O=, S= M3 = H, CH3, I, Br, CF3, NO2, OCH3, F, Cl, SCH3, SO2M; T2 = –O–, –NH– U2 = H, CH2, CH3, C6H5, etc; V2 = "–", CH2, CH3, C6H5, –CO– W2 = "–", CH2, CH3, –O–; X2 = "–", CH3, Br; and t2 = "–", H, CH3, C4H9.

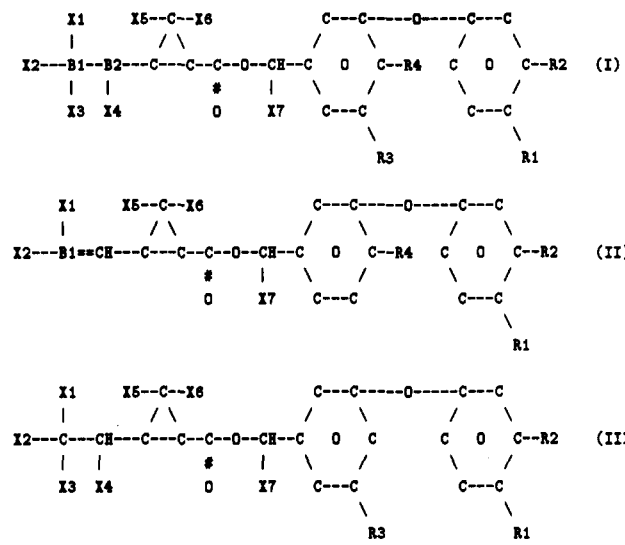


Figure 6. SFM of ether pyrethroid compounds. SFM(II) and SFM-(III) are derived from SFM(I).

Case Study 1: Sulfonyl Urea Herbicides with Benzoheterocycles. A data set of 205 compounds is taken from S–ADB. The original data source is from U.S. Patent 4854962[2] and U.S. Patent 4741761.[9] The definitions of frame site [A2], [a1], [MH], and [N1] are in ref 10. Partial data are shown in Table II. There are 11 grades of activities: 0, 1, 2, 3, ... 9, and A. Grade "A" means the activity is the highest, and grade "0" represents no activity. "/" indicates not tested. The method used in this case study is a combination of SFM and SDA. Using CASAC software, we have dealt with all 26 systems (13 kinds of target plants; for each plant there are two ways of testing—the postemergence and the preemergence). There are 15 frame sites defined by means of SFM. For every frame site, MR and EN parameters are used to form 30 independent variables. In selection of activity data, some medium activity compounds with activity grades from 1 to 8 are ignored to get a better data set. Table III, which originally displays on screen, shows the distribution of activity grades of all 205 compounds for 26 systems. Each row in Table III implies a natural profile of activities to a particular target plant, from which one can see most of compounds are at two ends of the profile, either high activity or low activity. In other words, the ordinary rule of classification, that we treat it as a two-class problem, is reasonable. For instance, at row 6 for barnyard grass (*Echinochloa crusgalli*) of

Table VI. Analysis of Important Frame Sites for Morningglory (*Ipomoea* sp.)[a]

F1	F2	D2	F(D2)	$N1_e$	$N2_e$	N_e	agreed %	N	no. and importance sequence of selected variables											
20	20	1.10	34.7	24	11	35	73.2	1	20											
15	15	4.44	27.1	10	6	16	87.7	5	20	5	8	16	19							
10	10	5.15	26.1	12	5	17	86.9	6	20	5	8	16	19	3						
8	8	5.15	26.1	12	5	17	86.9	6	20	5	8	16	19	3						
4	4	7.48	19.8	8	7	15	88.5	11	20	5	8	16	19	26	23	3	9	25	10	
2	2	7.70	18.5	6	9	15	88.5	12	20	5	16	8	19	27	9	23	25	3	10	2

[a] Mode of usage = postemergence; N_t = 130; $N1_t$ = 76; $N2_t$ = 54.

Table VII. Analysis of Important Frame Sites for Crabgrass (*Digitaria* sp.)

F1	F2	D2	F(D2)	$N1_e$	$N2_e$	N_e	agreed %	N	no. and importance sequence of selected variables									
20	20	2.71	92.0	9	15	24	82.4	1	20									
10	10	3.25	54.8	9	13	22	83.8	2	20	22								
9.6	9.6	3.78	42.1	9	11	20	85.3	3	20	22	15							
9.5	9.5	7.45	34.5	9	6	15	89.0	7	20	5	19	(21)	15	16	8	22		
9	9	7.45	34.5	9	6	15	89.0	7	20	5	19	(21)	15	16	8	22		
5	5	8.01	32.2	7	6	13	90.4	8	20	5	19	15	16	22	8	3		
3	3	8.01	32.2	7	6	13	90.4	8	20	5	19	15	16	22	8	3		
2	2	8.29	29.4	7	6	13	90.4	9	20	5	19	15	16	8	22	3	1	
1	1	10.24	19.3	6	7	13	90.4	16	20	5	19	15	22	16	24	10	8	...

[a] Mode of usage = postemergence; N_t = 136; $N1_t$ = 65; $N2_t$ = 71.

Table VIII. Analysis of Important Frame Sites for Wheat (*Triticum aestivum*)

F1	F2	D2	F(D2)	$N1_e$	$N2_e$	N_e	agreed %	N	no. and importance sequence of selected variables						
10	10	1.86	75.6	24	13	37	78.4	1	20						
9	9	1.86	75.6	24	13	37	78.4	1	20						
8	8	1.86	75.6	24	13	37	78.4	1	20						
7.2	7.2	1.86	75.6	24	13	37	78.4	1	20						
7	7	4.60	36.6	5	16	21	87.7	5	20	5	19	8	26		
5	5	4.60	36.6	5	16	21	87.7	5	20	5	19	8	26		
3	3	5.09	28.6	3	16	19	88.9	7	20	5	19	26	8	16	23
1.5	1.5	6.44	18.8	4	10	14	91.8	13	...						

[a] Mode of usage = postemergence; N_t = 171, $N1_t$ = 104, $N2_t$ = 67.

Table IX. Activity Prediction of CHA Compounds by NLM Method of CASAC

no. of group	left compds	activity class known	activity class predicted	no. of group	left compds	activity class known	activity class predicted
1	3	1	1	6	14	1	1
	10	1	1		38	1	1
	12	1	1		35	1	2
	43	1	1		45	2	2
2	2	2	2	7	11	1	1
	49	2	2		15	2	2
	20	2	2		17	1	1
	22	2	2		18	1	1
3	9	1	1	8	29	1	2
	16	1	1		30	1	2
	40	1	2		39	2	2
	42	1	1		41	2	2
4	1	1	1	9	24	1	1
	21	1	1		31	1	1
	32	1	2		44	1	1
	37	2	2		36	2	2
5	7	2	2	10	5	1	1
	4	1	1		6	1	2
	19	1	1		25	1	2
	33	2	2		13	2	2

[a] Class 1 = high activity. Class 2 = low activity. Seven of 40 compounds are misclassified. The rate of correct prediction is 82%.

postemergence, there are 112 high active compounds, 48 inactive compounds, and 45 medium ones.

In a series of computer-based experiments about classification, we obtain acceptable results. First of all, we pick out 10–12 main variables from the 30 independent variables with the SDA method. For instance, to barnyard grass (*Echinochloa crusgalli*) of postemergence, only seven compounds are misclassified within 160 samples. The rate of correct classification reaches 95.6%. All the classification results are summarized in Table IV. The average rate of correct classification of 26 systems is 92.0%.

Using the cross-validation (leave one out) method for activity prediction, the results are summarized in Table V, and they are also derived from 10–12 main variables. The average rate of correct prediction is 86.9% for 13 systems involving the postemergence treatment.

In the practical application of CASAC system, a generator of new compounds may produce thousands of "unknown" compounds, and then the CASAC system automatically predicts their bioactivities for many kinds of target plants, in this case 13, before the researchers synthesize any new compound. In this way, the CASAC is a helpful tool for synthetic chemists and may help people to save a lot of resources, funds, and time.

Another interesting result is that the system can help us to analyze the importance of frame sites in a SFM.

The automatic choice of important variables in an intrinsic attribute of the SDA method and other stepwise methods. In CASAC a frame site in a chemists' mind exactly is a mathematical variable. Variation of the F-factor's value in the run of SDA will lead to a change in the number of chosen variables. For studying which variables are important and what is the importance order, the authors have done three groups of computer testings. The chosen target plants are morning glory (*Ipomoea* sp.), crabgrass (*Digitaria* sp.), and

Table X. Input Data File of CASAC on 20 of 47 Ether Pyrethroid Insecticides[a]

	activity		frame sites												
cmpd	A1	A2	1	2	3	4	5	6	7	8	9	10	11	12	13
!---------	LC50	TL48	X1	X2	X3	X4	B1	B2	X5	X6	X7	R1	R2	R3	R4
001	0.01	>4.0	CL	CL	-	-	>C=	=CH-CL		CL	CN	H	H	H	H
002	0.01	3.8	BR	BR	-	-	>C=	=CH-CL		CL	CN	H	H	H	H
003	0.01	>4.0	CF3	CL	-	-	>C=	=CH-CL		CL	CN	H	H	H	H
004	0.02	>4.0	CF3	BR	-	-	>C=	=CH-CL		CL	CN	H	H	H	H
005	0.03	>4.0	CH3	CH3	-	-	>C=	=CH-CL		CL	CN	H	H	H	H
006	0.01	3.6	BR	CL	CL	BR	>C<	>CH-CL		CL	CN	H	H	H	H
007	0.01	3.5	BR	BR	CL	CL	>C<	>CH-CL		CL	CN	H	H	H	H
008	0.02	>4.0	CL	CL	CL	H	>C<	>CH-CL		CL	CN	H	H	H	H
009	0.02	>4.0	-CC-	-CC-		-	>C=	=CH-CL		CL	H	H	H	H	H
010	0.01	3.7	-TS-	-CC-		-	>C=	=CH-CL		CL	H	H	H	H	H
011	0.01	>4.0	OCH3-		-	-	-N=	=CH-CL		CL	CN	H	H	H	H
012	0.03	>4.0	OET	-	-	-	-N=	=CH-CL		CL	CH3	H	H	H	H
013	0.01	2.5	CL	CL	-	-	>C=	=CH-CF3		CL	CN	H	H	H	H
014	0.01	2.1	BR	BR	-	-	>C=	=CH-CF3		CL	CN	H	H	H	H
015	0.01	2.8	CF3	CL	-	-	>C=	=CH-CF3		CL	CN	H	H	H	H
016	0.02	3.7	CL	F	-	-	>C=	=CH-CF3		CL	CN	H	H	H	H
017	0.01	2.0	BR	CL	CL	BR	>C<	>CH-CF3		CL	CN	H	H	H	H
018	0.02	>4.0	BR	BR	H	CL	>C<	>CH-CF3		CL	CN	F	H	H	H
019	0.04	>4.0	F	F	F	H	>C<	>CH-CBR		F	CH3	H	CF3	CH3	H
020	0.04	>4.0	OPR	-	-	-	-N=	=CH-CH3		BR	C$C	OCH3	H	H	H

.....

[a] Compounds = 47; frame sites = 13; parameters taken = OI, OI^2, MR(X1, X2, X3, X4, B1, B2, X5, X6, X7, R1, R2, R3, R4), MR^2(X1, X2, X3, X4, B1, B2, X5, X6, X7, R1, R2, R3, R4), EN(X1, X2, X3, X4, B1, B2, X5, X6, X7, R1, R2, R3, R4), and EN^2(X1, X2, X3, X4, B1, B2, X5, X6, X7, R1, R2, R3, R4); variables = 2 + 13 × 4 = 54.

Table XI. QSAR Equations of Ether Pyrethroid Insecticides

system	N	QSAR equation	R	S	F1	F2	F3	F4	F5	F6	F7	F
(I)A	47											
A =		0.0291MR(R4) – 0.00094MR(X2) + 0.0095OI – 0.0026EN^2(X5) + 0.0162	0.8396	0.0067	19	17	13	12	6			4.00
(II)	33											
A =		–0.0035EN^2(X5) + 0.0203MR^2(R4) + 0.0011EN^2(R4) + 0.0018MR^2(X1) + 0.0382	0.8507	0.0066	18	15	14	5				4.00
(III)	14											
A =		–0.0192MR(X1) + 0.0066MR(X5) + 0.0006EN^2(R1) – 0.0036EN(X6) + 0.0402	0.9780	0.0036	42	19	13	5				4.00
(I)B	47											
A =		0.0020MR^2(X5) + 0.0304MR(R2) – 0.0890MR^2(R2) + 0.0059EN(R4) – 0.0355EN(X5) + 0.0045EN^2(X5) + 0.0119MR(R4) + 0.1250	0.8321	0.0070	11	9	6	6–	5+	4+	3	2.00

[a] Activity = LC_50 (ppm) for Akai Nika larva. For (I)B, New X1 = X1 + X2 + X3; new X5 = X5 + X6. F is F-Factor value assigned before a run. F1, F2, F3, F4, F5, F6, and F7 are F-test values calculated during a stepwise Hansch method running.

Table XII. Classification Results on 47 Ether Pyrethroid Insecticides

acty	class 1 (ppm)	class 2 (ppm)	total variable	selected variable	$N1_e/N1_t$	$N2_e/N2_t$	N_e/N_t	agreed %
LC$_50$	>=0.02	<0.02	54	2	9/25	2/22	11/47	76.6
Tlm48	>4.0	<4.0	54	7	5/26	4/21	9/47	80.9

[a] $N1_t$ is the total number of compounds in class 1. $N1_e$ is the number of misclassified compounds in class 1. $N2_t$ is the total number of compounds in class 2. $N2_e$ is the number of misclassified compounds in class 2. N_t is the total number of compounds. N_e is the number of misclassified compounds. In SDA, F1 = F2 = 2.0.

wheat (*Triticum aestivum*). When the F-factor value is large, which means that the criterion is strict, the number of selected variables should be reduced. Tables VI–VIII represent the testing results of the three systems with the same abbreviations as Table V. See Table VI. When the F-factor is assigned to be 20, only one variable (number 20) is selected. It is the EN parameter of frame site M2 (EN(M2)). When the F-factor is 15, five variables (number 20, 5, 8, 16, and 19) are chosen. They are EN(M2), respectively. In an overview of Tables VI–VIII, we can see that the variable 20 (EN(M2)) is the

most important one in every system. The next important groups are variables 5, 8, 19 (MR(M2), MR(M5), En(M1)), and so on.

Some important variables only appear in a special system, for example, variable 22 (EN(M4)) in crabgrass (*Digitaria* sp.)(see Table VII). It implies that there might be a special mechanism about the interaction or metabolic process, which is concerned with structural factors at those important frame sites. It should be stressed that in a biosystem, the interaction mechanism is very complicated. However, the result proposed by CASAC software is an enlightenment to deepen this knowledge. In fact, to all tested systems the frame site M2 (its MR and EN property) is the most noteworthy.

In summary, CASAC software can provide not only valuable statistical information about bioactivity of unknown compounds but also some information about specific characteristics of a biosystem.

Case Study 2: Chemical Hybridizing Agents. A set of chemical hybridizing agents (CHAs) (1-aryl-1,4-dihydro-4-oxo(thio)pyridazine derivatives) is taken as the study system.[11] The data set is from U.S. Patent 4345932[12] EP Patent 0138662,[13] and EP Patent 0137133.[14] It contains 51 compounds in all. The SFM is shown in Figure 5 and has 11 frame sites.

In this study, the NLM method in CASAC has been used with the cross-validation (leave four out) method. We have done the predictions 10 times, and each time we randomly pick out four compounds as "unknown" compounds from the 51 compounds. The correct rate of classification and prediction is 82%. The result is shown in Table IX. It indicates that the predicting ability is acceptable to design and screen new CHAs.

Case Study 3: Ether Pyrethroid Insecticides. This example refers not only to the qualitative classification problem (SAR) but also to the quantitative equation (QSAR). Furthermore, it is a relative two-targets problem. The two targets are insecticidal activity LC_{50} to *Akai Nika larva* (ppm) and TLm48 (ppm) toxic to fish. We need an insecticide which kills insects and has no or minor effect on fish. The activity data are from Japanese Patent 58-79909[15] and with which an input file of CASAC is formed as Table X. The SFM frames of 20 of the 47 ether pyrethroid compounds are shown in Figure 6. The SFM(I) in Figure 6 is divided into SFM(II) and SFM(III), because we want to know how the bond between B1 and B2 affects the activity. The Hansch QSAR equations are listed in Table XI, and SAR results are summarized in Table XII.

Summarizing the whole results of CASAC for this system makes us come to several conclusions, which may give the related chemists some helpful suggestions for their synthesis:

(a) The most important structural factors are substituents at frame site X5 and X6. The smaller the MR values are, the better the activity is, and the bigger the EN values are, the better the activity is. As a matter of course, the tendency is not unlimited.

(b) The second important structural factor is substituents at site R4. It is suggested that both MR and EN of substituents at site R4 be small.

(c) The third factor to be considered involves substituents at site X1, X2, and X3. It is worthy to have a try with bigger MR and smaller OI at those sites.

(d) It seems that a compound with SFM(III), a single bond between sites B1 and B2, may have higher activity to target *Akai Nika larva*.

The above results are in good accord with the results of Bushell,[16] who got his results from experience.

DISCUSSION

In summary, the SFM method and the CASAC system have some obvious advantages, particularly the practicality to improve screening processes. However, for efficiency, there is a requirement about the experimental data of biotesting—to rationally design a initial set of compounds. Both activity ranks and structural changes (substituents) are better with a wider profile. In our experience, the qualitative data are good enough to use CASAC, as long as the data profile is proper. The methodology of CAMD in LCC is to provide useful information as much as possible in each synthesis–biotesting–prediction–synthesis cycle.

We would like to indicate that any computer-aided molecular design software tool, including the CASAC system, should be combined with a comprehensive investigation of reaction mechanisms in molecule, cell, tissue, organ, and entirety levels. We believe that the common efforts coming from synthesis-chemists and computer-chemists will bring great benefits and the new development of pesticide chemistry.

In the next stage of the project, we will continue the CAMD study in two directions: (a) Put the three-dimensional and electronic structure information into the CASAC system, (b) A fundamental study of computer chemistry in heterocyclic structure chemistry.

ACKNOWLEDGMENT

The authors would like to thank Professor Y. Wolman, Department of Organic Chemistry, The Hebrew University of Jerusalem, Israel, for discussing the research direction of CAMD last year, encouraging, and supporting J.Z. to bring the paper to the 10th ICCCRE Conference. Financial support to the project from the Ministry of Chemical Engineering Industry and partly from the Natural Science Foundation of China and the Center for Scientific Databases, Chinese Academy of Sciences is gratefully acknowledged. We also thank Prof. Tu Junli, the Institute of Chengdu Organic Chemistry, Chinese Academy of Sciences, for introducing the activity data of the ether pyrethroid insecticides. Also we thank Prof. Yang Zhangyuan, Prof. Dong Qian, Prof. Yuan Shengang, Prof. Li Renli, Prof. Xu Lu, Dr. Yan Xinjian, and Dr. Zhang Yuemin for their beneficial discussions during the last years.

REFERENCES AND NOTES

(1) Hansch, C.; Leo, A. J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley: New York, 1979.
(2) Douglass, D. L.; Moon, M. P. U.S. Patent 4854962, 1989.
(3) Yuan, L.; Ding, Y. Organic Property Concept and Its Application in Surface Chemistry. *Huaxue Tongbao* **1988**, *2*, 19–21 (in Chinese).
(4) Note about OI parameter. The summary of the study is as follows: System 1, 10 kojic acid and its derivatives, $R = 0.979$, $S = 0.158$; System 2, 22 *N*-chloroacetyl-*N*-(2,6-diethylbenzene)aminoethyl ester compounds, $R = 0.955$, $S = 0.332$; System 3, 37 *N*-chloroacetyl-*N*-phenylglycine ester, $R = 0.955$, $S = 0.198$; System 4, 9 derivatives of β-substitute cassia bark amides, $R = 0.957$, $S = 0.239$; System 5, 28 erythromycin esters, $R = 0.978$, $S = 0.076$; System 6, 6 sulfonylurea compounds, $R = 0.905$, $S = 0.493$; System 7, 23 dihydrosafrole metabolite, $R = 0.926$, $S = 0.339$; System 8, 18 aryltriazenes, $R = 0.949$, $S = 0.307$; System 9, 40 dipeptides or tripeptides, $R = 0.891$, $S = 0.367$; System 10, 53 *N*-acetyl dipeptide amides or *N*-acetyl tripeptide a, $R = 0.905$, $S = 0.352$.
(5) Han, C. The Calculation of the Ionic Group Electronegativities and Neutral Group Electronegativities. *Acta Chim. Sin.* **1990**, *48*, 627–31.
(6) Note about EN parameter. EN is defined as the average of energy over the efficient number of valence electrons of center atom. Group EN is concerned with the charge and lone pair of center atom, bonds, and p–π conjugation between the center atom and its immediately connected atoms or groups, and group EN's atoms or groups that center atom immediately connects to. After self-consistence, we can easily know the charge distribution in a group or molecule. The details will be published elsewhere.

(7) Probability Group in Computer Center, Chinese Academy of Sciences. *A Package for Statistical Analysis of Stochastic Data*; User's Guide in Chinese; SASD: Beijing, China, 1987.
(8) Sun, D.; Xie, G.; Yan, X.; Wang, H.; Zhang, Y.; Cao, I.; Zhou, J.; Xu, Z. In *Selected Papers of Computers and Applied Chemistry (1991)*, Computers Chemistry Monograph Series 2; Editorial Board of Computers and Applied Chemistry & LCC, Chinese Academy of Sciences Ed.; Science Press: Beijing, China; 1991, p 61–67.
(9) Pasteris, R. J. U.S. 4741761, 1986.
(10) Note: [A1] and [A2] form the main chain. [a1] is a branch on [A1]. [N1] is defined as a substituent on atom N. [MH] is a special frame site to determine that the benzo-ring contains 5 or 6 members.
(11) Zhou, J.; Yan, X.; Sun, D.; Cao, L.; Xu, Z.; Chen, W. A 2D-Structure–Selective Male Sterility Relationship Study on 1-Aryl-1,4-Dihydro-4-Oxo(Thio)Pyridazine Derivatives. *Acta Agric. Univer. Pekinensis* **1991**, *17*, 2, 11–17. (in Chinese).
(12) Fujimoto, T. T. U.S. Patent 4345932, 1982.
(13) Labovitz, J. N.; Fang, L. EP Patent 0138662, 1985.
(14) Haddock, E.; Hopwood, W. J. EP Patent 0137133, 1981.
(15) Fujita, J. Jpn. Patent 58-79909, 1983.
(16) Bushell, M. J. *Recent Advanced in the Chemistry of Insect Control II*; 1990; p 125–141.