COMPUTER-AIDED PREDICTION OF ACTIVITY SPECTRUM FOR SUBSTANCE (PASS) SYSTEM EVALUATED ON A SET OF NEW BIOLOGICALLY ACTIVE COMPOUNDS

T. A. Gloriozova,[1] D. A. Filimonov,[1] A. A. Lagunin,[1] and V. V. Poroikov[1]

Translated from Khimiko-Farmatsevticheskii Zhurnal, Vol. 32, No. 12, pp. 33 – 39, December, 1998.

Original article submitted February 24, 1998.

The research and synthetic work necessary for the creation of new drug preparations requires considerable expenditures and involves a risk of obtaining negative results because of unpredicted side effects and toxicity. The use of computer-aided predictions of the main and side effects of a pharmacologically active compound in the early stage of its characterization may optimize the procedure of selecting biological structures for more profound study and reduce the costs of research and development (R & D) work.

The computer-aided system of Prediction of Activity Spectrum for Substance (PASS, Versions 3.20 – 4.30) predicts 114 types of biological activity based on the structural formula of a given compound, including the pharmacological effects and mechanisms of action [1 – 3]. It was demonstrated that the cfficiency of application of this computer-aided approach in the screening of new compounds exceeds 500%. The accuracy of the computer-aided prognosis exceeds the accuracy of experts' predictions by 300% [3]. Recently, the predicting ability of the PASS system has been extended to 414 types of biological activity, the new version called PASS-C. The mean accuracy of PASS-C predictions with a sliding quality control (leave-each-out cross-validation) procedure is 83.4%. However, this cross-validation method is known to provide too optimistic estimates in the case of degeneracy of the learning set. Therefore, it was of interest to evaluate the PASS-C system under rigid conditions of predicting biological activity using a set of new (recently reported) biologically active compounds.

METHOD OF BIOLOGICAL ACTIVITY PREDICTION

The methods of prediction employed in the PASS system have been recently described in [1, 2]. The PASS-C version has been significantly improved: (1) the list of predictable activities is extended to 414 types; (2) the chemical structures are described within a new approach; (3) the mathematical procedures are refined. The principles of the modified method are outlined below.

Description of the Structure of Chemical Compounds

In the existing PASS system, the structures of chemical compounds were described based on their two-dimensional structural formulas. These formulas can be read, for example, with the aid of a special editor ISIS/Draw (MDL Information Systems Inc.). The standard representation of structural data for the computer consists of *.mol files for a single molecule and *.sdf files for a set of compounds.

Upon the subsequent computer processing, the structural formula is represented by a list of atoms forming the given molecule and a list of bonds between these atoms. On this ground, a set of structure descriptors, or a basis set of atomic environments, is generated for each chemical compound. For this purpose, the list of atoms is complemented by all the necessary hydrogen atoms attached in accordance with the valence rules. Then the cycles are revealed and the atoms forming these cycles are designated. Finally, a descriptor of environment is recursively constructed for each atom as described below.

All elements of the Periodic Table are subdivided into several classes (in PASS-C, there are 32 such classes) taking into account their role in biological activity. For example, all lanthanides can be placed in the same class (bioisosterism). The descriptor of atomic environment of a given atom includes (i) the number of the class to which this atom belongs; (ii) a mark indicating whether this atoms enters into a cycle; and (iii) descriptors of the nearest neighbors (in lexicographic order), that is, atoms linked by chemical bonds to the given atom. In the first step of the recursive procedure, the descriptors include only the numbers of classes and the marks of entering into cycles. As a result, each atom is characterized by a line of symbols describing its environment. Our experiments showed that an optimum recursive procedure to recognize a biological activity consists of two steps, corresponding to the environment including all the neighboring atoms at a distance of one (first order) and two (second order) bond lengths. The

[1] Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, Russia.

0091-150X/98/3212-0658$20.00

TABLE 1. Characteristics of the Learning Set of the PASS System and Prognosis Quality for the Leave-Each-Out Cross-Validation

Activity type	Number of substances	E1, %	E2, %	DE, %	AvE, %
Antitumor	2672	18.34	19.05	0.71	18.69
Antihypertensive	2545	21.38	21.61	0.24	21.49
Antibacterial	1308	22.17	13.63	8.54	17.90
Psychotropic	1265	13.68	23.09	9.42	18.38
Antiallergic	1231	17.22	26.63	9.41	21.93
Spasmolytic	1180	12.54	29.32	16.78	20.93
Antiviral	930	29.89	13.85	16.05	21.87
Antifungal	907	26.13	13.66	12.47	19.90
Cognitive function activation	874	19.11	22.07	2.97	20.59
Antiinflammatory	796	21.61	22.66	1.05	22.13
Antidepressant	791	11.00	27.99	16.99	19.50
Anticonvulsant	751	18.51	24.35	5.84	21.43
Hypolipidemic	733	11.05	15,,39	4.34	13.22
Antiinflammatory (nonsteroidal)	698	14.61	24.28	9.67	19.45
Thrombocyte aggregation inhibition	687	14.27	21.56	7.30	17.91
Vasodilative	663	17.20	31.14	13.95	24.17
Analgesic	658	10.64	24.44	13.80	17.54
...	...	...	...	...	...

descriptors of environment are constructed so as to retain the integrity of a molecular fragment, while remaining single-valued, whereby equivalent fragments have identical descriptors.

For organic compounds containing a comparatively large number of tetravalent carbon atoms, second-order descriptors correspond to whole molecular fragments including 5–17 atoms, while third-order descriptors would involve 8–53 atoms. The lines of symbols, representing descriptors of the environment of each atom, form a vocabulary which constitutes a basis set of environments of a given molecule. A description of the molecular structure represents a list of descriptor numbers according to this vocabulary, which characterizes the basis set of environments.

Computer simulations showed that descriptors of third-order environments are practically unique and the size of the vocabulary is enormously large. The supplement of a second-order basis set of environments by the first-order basis set (without taking into account the possibility of atoms entering into cycles) insignificantly increases the vocabulary volume, while providing a considerably improved quality of prognosis. This variant has proved to be optimum among all the possible basis set combinations.

Algorithms of Prediction and Prognosis Quality Control

As described above, each compound of the learning set is represented in the PASS system by a list of numbers of the environment descriptors and a list of numbers corresponding to biological activities exhibited by this compound.

The process of predicting the biological activity spectrum of a chemical compound is as follows.

First, the structural formula of a new (predicted) compound is transformed into the corresponding environment basis set and the corresponding descriptor numbers are found in the vocabulary of environment descriptors. Some descriptors may be not found in the vocabulary (the corresponding fragments are present in no one molecule of the learning set). If all the descriptors are of this type, the compound has no common fragments with compounds in the learning set and its biological activity spectrum cannot be predicted by the PASS system. When the number of such new descriptors is rather large, the results of prognosis should be considered only as orientational.

Using the numbers of descriptors of the environment basis set, the system determines the following quantities, which are calculated using the data available for the learning set:

n, the number of compounds in the learning set;

n_i, the number of compounds for which the environment basis set contains the i th descriptor;

n_j, the number of compounds possessing the j th activity;

n_{ij}, the number of compounds having an environment basis set containing the i th descriptor and possessing the j th activity;

$V = \Sigma n_j$, the total number of environment descriptors in the learning set;

$V_j = \Sigma n_{ij}$, the total number of environment descriptors in compounds possessing the j th activity.

First, the system calculates estimates of the probability of manifestation of the j th activity provided the i th environment descriptor is present,

$$P_{ij} = \frac{n_{ij}}{n_i},$$

or estimates of the probability of manifestation of this activity in the absence of any significant relationship between descriptors and the j th activity:

$$P_j = \frac{V_j}{V}.$$

Then, the descriptors found in the vocabulary for the environment basis set of the predicted molecule are used to calculate the following sum:

$$t_j = \sum \left\{ Arcsin(a(2P_{ij} - 1)) - Arcsin(a(2P_j - 1)) \right\},$$

where a is the correcting coefficient (depending on n_i and the dimension of the environment basis set of the predicted mole-

TABLE 2. PASS Prognosis of Activities for Substances of the Evaluation Set

Ref.	Compound	Activity	Prognosis / experiment
[4]	Ia	Anticonvulsant	+/+
	Ib	Anticonvulsant	+/+
	Ic	Anticonvulsant	+/+
[5]	III	Diuretic	+/+
		Saluretic	+/+
	IV	Diuretic	+/+
		Saluretic	+/+
	V	Diuretic	+/+
		Saluretic	+/+
	VI	Diuretic	+/+
		Saluretic	+/+
	VII	Diuretic	+/+
		Saluretic	+/+
[6]	VIII	Antiviral	–/+
	XI	Antiviral	–/+
	XII	Antiviral	–/+
	XIII	Antiviral	–/+
[7]	V	Anticholinesterase	–/+
	VI	Anticholinesterase	–/+
	VII	Anticholinesterase	+/+
	VIII	Anticholinesterase	+/+
	IX	Anticholinesterase	–/+
	X	Anticholinesterase	–/+
	XI	Anticholinesterase	–/+
	XII	Anticholinesterase	–/+
	XIII	Anticholinesterase	+/+
	XIV	Anticholinesterase	–/+
	XV	Anticholinesterase	+/+
	XVI	Anticholinesterase	–/+
	XVII	Anticholinesterase	+/+
	XVIII	Anticholinesterase	+/+
	XIX	Anticholinesterase	+/+
[8]	I	Spasmolytic	+/+
		Cholagogic	–/+
	II	Spasmolytic	+/+
		Cholagogic	–/+
	III	Spasmolytic	+/+
		Cholagogic	+/+
	IV	Spasmolytic	+/+
		Cholagogic	+/+
	V	Spasmolytic	+/+
		Cholagogic	–/+
[9]	I	Antiinflammatory	+/+
	II	Antiinflammatory	–/+
	IX	Antiinflammatory	+/+
	XII	Antiinflammatory	–/+
	XIII	Antiinflammatory	+/+
[10]	I	Anticonvulsant	+/+
	II	Anticonvulsant	+/+
	III	Anticonvulsant	+/+
	V	Anticonvulsant	+/+
[11]	II	Antihypertensive	-/+
	III	Antihypertensive	+/+
	IV	Antihypertensive	+/+
	V	Antihypertensive	+/+
	VI	Antihypertensive	+/+
	VII	Antihypertensive	+/+
	VIII	Antihypertensive	+/+
	IX	Antihypertensive	+/+
[12]	VIIIa	M-cholinoblocking	+/+
	VIIIb	M-cholinoblocking	+/+
	VIIIc	M-cholinoblocking	–/+
	VIIIe	M-cholinoblocking	+/+
	VIIIf	M-cholinoblocking	+/+
	VIIIg	M-cholinoblocking	+/+
[13]	VI	Immunotropic	+/+
		Antiinflammatory	+/+
	VII	Immunotropic	+/+
		Antiinflammatory	–/+
	VIII	Immunotropic	+/+
		Antiinflammatory	+/+
	IX	Immunotropic	+/+
		Antiinflammatory	–/+
	X	Immunotropic	+/+
		Antiinflammatory	+/+
	XI	Immunotropic	+/+
		Antiinflammatory	–/+
[14]	IId	Antibacterial	+/+
	IIe	Antibacterial	+/+
	IIf	Antibacterial	+/+
	IIg	Antibacterial	+/+
	IIh	Antibacterial	+/+
	IIi	Antibacterial	+/+
[15]	IVa	Analgesic	–/+
		Anticonvulsant	–/+
	IVb	Analgesic	+/+
		Antiinflammatory	+/+
	IVc	Analgesic	+/+
		Antiinflammatory	+/+
	IVd	Analgesic	–/+
	IVe	Analgesic	+/+
	IVf	Analgesic	+/+
		Antiinflammatory	+/+
[16]	I	Anticholinesterase	+/+
		Potassium antagonist	+/+
	II	Anticholinesterase	+/+
		Potassium antagonist	–/+
	III	Anticholinesterase	+/+
		Potassium antagonist	+/+
	VI	Anticholinesterase	+/+
		Potassium antagonist	+/+
	V	Anticholinesterase	+/+
		Potassium antagonist	–/+
[17]	IIIa	Antiarrhythmic	–/+
		Antiaggregant	+/+
	IIIb	Antiarrhythmic	–/+

TABLE 2. (Continued)

Ref.	Compound	Activity	Prognosis / experiment	Ref.	Compound	Activity	Prognosis / experiment
		Antiaggregant	+/+		IVg	Antiarrhythmic	+/+
	IIIc	Antiarrhythmic	–/+			Antioxidant	–/+
		Antiaggregant	+/+			Hypoglycemic	–/+
	IIId	Antiarrhythmic	–/+		IVh	Antiarrhythmic	+/+
		Antiaggregant	+/+			Antioxidant	–/+
	IIIe	Antiarrhythmic	–/+			Hypoglycemic	–/+
	IIIf	Antiarrhythmic	–/+		IVi	Antiarrhythmic	–/+
	IIIh	Antiarrhythmic	–/+			Antioxidant	+/+
[18]	IVa	Antihypertensive	+/+		IVj	Antioxidant	+/+
		Antiarrhythmic	+/+		IVk	Antioxidant	+/+
		Antioxidant	–/+			Hypoglycemic	+/+
		Hypoglycemic	–/+		IVl	Antihypertensive	+/+
	IVb	Antiarrhythmic	+/+			Antioxidant	–/+
		Antioxidant	–/+		IVm	Antioxidant	+/+
		Hypoglycemic	–/+			Hypoglycemic	–/+
	IVc	Antioxidant	–/+		IVn	Antioxidant	+/+
	IVd	Antiarrhythmic	+/+		IVo	Antiarrhythmic	–/+
		Antioxidant	–/+			Antioxidant	–/+
		Hypoglycemic	–/+		IVp	Antioxidant	+/+
	IVe	Antioxidant	+/+		IVq	Antihypertensive	+/+
		Hypoglycemic	–/+			Antioxidant	–/+
	IVf	Antihypertensive	+/+		IVr	Antioxidant	+/+
		Antioxidant	–/+				

cule) selected so as to provide the best quality of predictions based on the given learning set.

Another sum is calculated by the formula

$$d = \sum \frac{1}{n_i + 1.216}$$

and gives an estimate of dispersion of the t_j quantity.

In the next step, each predicted activity type is characterized by estimated probability for the given compound to belong to a subset of chemical compounds of the learning set possessing the given activity type:

$$P_j = \frac{1}{1 + Exp\left(-2t_j \cdot \frac{t_j^2}{t_j^2 + d}\right)}.$$

All estimates calculated as described above are arranged in decreasing order. The predicted biological activity spectrum is presented in the form of a list of activities characterized by $P_j > 1/2$.

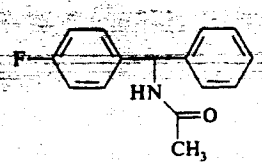
The quality of prognosis is assessed using a sliding control procedure (leave-each-out cross-validation). According to this, each compound from the learning set is characterized by the quantities t_j calculated for the n, n_i, n_j, n_{ij}, V, V_j values determined upon excluding this compound from the learning set. Then, provided the given compound has the j th activity and $t_j < 0$, we obtain an error of the first kind (active substance classified as inactive). If the compound does not possess the j th activity and $t_j > 0$, we have an error of the second kind (inactive substance classified as active). The quality of the whole prognosis is characterized by the average probabilities of errors of the first (E1) and second kind (E2), the average modulus of difference between probabilities of the two kinds (DE), and the average of the probabilities of errors of the two kinds (AvE).

Description of Biological Activity, Characteristics of the Learning Set, Cross-Validation Results

The biological activities are assessed in the PASS system on a qualitative level (yes / no). Each compound in the learning set is provided with a list of numbers of the biological activity types found in this compound. In forming the list of predictable activities, we proceeded from a criterion widely used in the analysis of qualitative structure – activity relationships (QSAR). According to this, the number of chemical compounds representing a given parameter (in this case, the activity type) must be not less than five. Table 1 shows some types of activity predicted by the PASS-C system, indicating the number of substances in the learning set possessing every type of activity and the characteristic of prognosis quality for the leave-each-out cross-validation.

The complete list of biological activity types that can be predicted by the PASS-C1 system, with the cross-validation estimates of the prediction quality, is available on the Internet from http://www.ibmh.msk.su / PASS/.

TABLE 4. Example of Prediction of the Biological Activity Spectrum for a Compound [10]*:



Activity	Probability of manifestation	Activity	Probability of manifestation
Anticonvulsant	0.999	Platelet activating factor antagonist	0.560
Acetyl CoA transferase inhibitor	0.972	Sympatholytic	0.560
Prostate disorders treatment	0.937	Benzodiazepine agonist, partial	0.559
5α-Reductase inhibitor	0.924	Mucolytic agent	0.559
Androgen antagonist	0.857	Protease inhibitor	0.555
Oxazolidinone-like antibiotic	0.852	Histamine H2 receptor antagonist	0.555
Antianginal agent	0.827	Protocollagen prolyl hydroxylase inhibitor	0.553
Alpha 1 adrenoreceptor antagonist	0.812	Neuropeptide Y antagonist	0.553
Thrombin inhibitor	0.789	GABA receptor agonist	0.551
Luteinizing hormone-releasing hormone antagonist	0.785	Adrenergic transmitter uptake inhibitor	0.551
5-Hydroxytryptamine-2 agonist	0.771	5 Hydroxytryptamine reuptake inhibitor	0.551
Respiratory analeptic	0.735	Analgesic, opioid	0.550
Platelet aggregation inhibitor	0.728	Nitric oxide donor	0.548
Growth hormone release promoting agent	0.715	Immunomodulator (AIDS)	0.545
Hypolipemic	0.715	Dermatologic	0.543
Urinary incontinence treatment	0.704	Skeletal muscle relaxant	0.542
Antihypoxic	0.698	Microtubule formation inhibitor	0.537
Melatonin agonist	0.690	MAO-B inhibitor	0.537
Anticoagulant	0.679	Psychostimulant	0.535
Analeptic	0.665	GABA A receptor agonist	0.534
Interleukin 1 beta converting enzyme inhibitor	0.661	Opioid mu receptor agonist	0.533
Hair growth promoter	0.653	Aldose reductase inhibitor	0.532
Dopamine antagonist	0.646	Dopamine uptake inhibitor	0.532
Dopamine beta hydroxylase inhibitor	0.644	Complement inhibitor	0.528
Neurokinin antagonist	0.638	Liver fibrosis treatment	0.527
Substance P antagonist	0.633	Catechol O methyltransferase inhibitor	0.526
Phospholipase A2 inhibitor	0.632	Dopamine D1 antagonist	0.524
Antidyskinetic	0.627	Hematopoietic	0.523
Lipoxygenase inhibitor	0.623	Vasopressin 1 antagonist	0.517
Neuroleptic	0.620	Antiinfective (AIDS)	0.517
Sigma antagonist	0.617	Endothelia formation inhibitor	0.516
Anorexic	0.613	Anticataract	0.516
Opioid delta receptor agonist	0.606	Matrix metalloproteinase inhibitor	0.513
Calcium channel antagonist	0.604	Aminopeptidase microsomal inhibitor	0.512
Melatonin antagonist	0.603	Anti-HIV agent	0.511
Antipsychotic	0.596	Prostaglandin E2 antagonist	0.509
Hypoglycemic	0.586	Psychosexual dysfunction treatment	0.509
Ribonucleotide reductase inhibitor	0.579	Antiparkinsonian	0.507
Dopamine D2 antagonist	0.578	5 Hydroxytryptamine agonist	0.506
Antihistaminic	0.569	Uricosuric	0.505
Antiacne	0.565	Corneal wound healing stimulator	0.504
Expectorant	0.563	MAO inhibitor, irreversible	0.503
GP IIb / IIIa antagonist	0.562		

* The structure contains 15 descriptors, which are all present in the learning set. PASS predicts 85 activity types (of 414); the anticonvulsant activity reported in the literature has a probability of 99.9%.

As seen from Table 1, the types of activity most widely represented in the learning set are antitumor, antihypertensive, antibacterial, psychotropic, and antiallergic. The best characteristics of the prognosis quality according to the leave-each-out cross-validation are obtained for antibiotics of different groups (cephalosporins, quinolones, taxanes, naphthyridines, etc.), carboanhydrase inhibitors, motilin analogs, vitamin D analogs, and glucocorticoid agonists. The worst quality characteristics are obtained for myorelaxants (skeletal muscles), dermatological preparations, complement inhibitors, and MAO inhibitors. In the general case, the more specific a particular pharmacological effect is and the less similar are the structures of substances possessing this activity to those of representatives of the other groups, the higher the quality of prognosis (and vice versa). As noted above, the average accuracy of predictions for the leave-each-out cross-validation is 83.4%, which allows the PASS-C system to be used in practice.

Principles of Evaluation Set Formation and Prediction of Activity

As pointed out above, the most rigid testing of the prognosis system is achieved in the case when the evaluation set contains substances possessing a certain novelty with respect to those entering into the learning set. The learning set of the PASS-C system includes substances representing either parent compounds of the drugs already used in medicinal practice or the compounds patented abroad over 1987 – 1996.

We have formed a new evaluation set based on the biologically active substances reported in the *Khimiko-Farmatsevticheskii Zhurnal* (*Pharmaceutical Chemistry Journal*) during the past two years (1996 – 1997), whose structures do not enter into the existing learning set of the PASS-C system. In order to provide for the generality of evaluation, the testing set must be sufficiently "heterogeneous" with respect to both the chemical classes of compounds and the activity types. Taking this point into account, we have taken 103 compounds representing various chemical classes, possessing experimentally verified biological activity of 18 types (Table 2; the names of activities and the numbers of substances are given according to the original publications).

It should be noted that an additional "rigidity" to this evaluation is imparted by the large variety (unavoidable in using the published data) of methods used for the biological activity determination. The same term (activity type) is frequently characterized by different experimental parameters. For this reason, the PASS system predicts activities on the qualitative level (yes / no); this circumstance is one of the factors decreasing the accuracy of predictions.

Table 2 shows the results of predicting the biological activity of compounds in the evaluation set by the PASS-C system. Here, the +/+ symbol indicates the presence of activity in both prognosis and experiment and –/+ corresponds to activity absent in the prognosis but found in experiment (error of the first kind). Because publications usually present data mostly for active compounds, it was impossible to estimate the average accuracy of predictions for the absence of activity (–/–) or the presence of errors of the second kind (+/–).

The evaluation set contained a total of 103 compounds possessing 151 types of activity. The PASS-C system correctly predicted 98 cases of activity (64.9%), errors of the first kind amounting to 35.1%.

Let us consider the expected frequency of accidentally "guessing" the known activity in this experiment. The PASS-C system predicts 414 activity types. The probability of an accidental coincidence of the known activity in predicting a single activity type (of 414) is therefore $1 / 414 = 0.0024$. Actually, each compound of the evaluation set was characterized by (on the average) 86 probable activities (Table 3). On the whole, 8900 activities were predicted for 103 compounds, which increases the probability of accidental coincidence to $(8900 \times 0.0024)/151 = 0.14$ (or 14%). Thus, the accuracy of prognosis with the PASS system is approximately $64.9 / 14 = 4.6$ times higher than the probability of accidental correct "guessing" of the activity. Taking into account the degree of novelty of the evaluation set as compared to the learning set, this result seems to be quite good.

Let us consider the quality of prognosis for each particular activity type. As seen from the data presented in Table 4, the accuracy of predictions exhibits a considerable scatter. The best results were obtained in predicting diuretic, saluretic, spasmolytic, immunotropic, antibacterial, and antiaggregant activities (100%). The worst predictions were made with respect to antiviral, hypoglycemic, and antiarrhythmic properties. The unsatisfactory results obtained for these activity types point to the necessity of re-filling the learning set of

TABLE 3. Prognosis Accuracy for Some Activities of Compounds in the Evaluation Set

Activity	Number of substances possessing this activity	Number of correct predictions	Errors of 1st kind, %
Anticonvulsant	8	7	12.5
Diuretic	5	5	0
Saluretic	5	5	0
Antiviral	4	0	100
Anticholinesterase	20	12	40
Spasmolytic	5	5	0
Cholagogic	5	2	60
Antiinflammatory	14	9	35.7
Antihypertensive	12	11	8.3
M-cholinoblocking	6	5	16.7
Immunotropic	6	6	0
Antibacterial	6	6	0
Analgesic	6	4	33.3
Potassium antagonist	5	3	40.0
Antiarrhythmic	14	5	64.3
Antiaggregant	4	4	0
Antioxidant	18	8	55.6
Hypoglycemic	8	1	87.5

the PASS system with the corresponding chemical compounds. In the case of antiviral activity, this extension must involve sulfamoylphenyl derivatives of succinamic acids; the hypoglycemic activity class should be filled with 4-dialkylaminoethylpyrrolo[1,2-a]benzimidazole derivatives; the antiarrhythmic activity group should include some amidino acids of the 3,4-dihydroisoquinoline series.

CONCLUSIONS

As seen from the results of evaluation of the PASS-C system performed with an independent set of compounds, the accuracy of prognosis is 4.6 times the probability of accidental coincidence. This implies that use of the PASS system in planning the syntheses of compounds with required properties and / or determining the tests relevant to particular substances would provide an economic efficiency of about 460%. It is interesting to note that this efficiency level is close to the value calculated previously for a heterogeneous set of 50 substances [19], which amounted to about 500% [20].

ACKNOWLEDGMENTS

The authors are grateful to MDL Information Systems, Inc., for kindly providing the ISIS program package and data bases for this investigation.

REFERENCES

1. D. A. Filimonov, V. V. Poroikov, E. I. Karaicheva, et al., *Eksp. Klin. Farmakol.*, **58**(2), 56 – 62 (1995).
2. D. A. Filimonov and V. V. Poroikov, *Bioactive Compound Design: Possibilities for Industrial Use*, BIOS Scientific Publ., Oxford (1996), pp. 47 – 56.
3. V. V. Poroikov, D. A. Filimonov, A. V. Stepanchikova, et al., *Khim.-Farm. Zh.*, **30**(9), 20 – 23 (1996).
4. A. I. Bokanov, M. I. Evstratova, K. F. Turchin, et al., *Khim.-Farm. Zh.*, **31**(10), 27 – 29 (1997).
5. V. A. Smirnov and L. V. Simerzina, *Khim.-Farm. Zh.*, **31**(10), 37 – 38 (1997).
6. D. Burdulene, Z. Stumbryavichute, Z. Talaikite, et al., *Khim.-Farm. Zh.*, **31**(9), 24 – 26 (1997).
7. T. G. Nikolaeva, A. A. Safonova, and A. P. Karpenko, *Khim.-Farm. Zh.*, **31**(9), 27 – 29 (1997).
8. A. N. Nabiev and V. I. Vinogradova, *Khim.-Farm. Zh.*, **31**(9), 30 – 32 (1997).
9. G. A. Tolstikov, V. P. Krivonogov, B. I. Galimov, et al., *Khim.-Farm. Zh.*, **31**(8), 26 – 29 (1997).
10. V. A. Krasnov, V. K. Gorshkova, A. A. Bakibaev, et al., *Khim.-Farm. Zh.*, **31**(7), 35 – 36 (1997).
11. E. G. Belozertseva, B. A. Chakchir, O. V. Solod, et al., *Khim.-Farm. Zh.*, **31**(6), 14 – 16 (1997).
12. A. Ya. Bespalov, T. L. Gorchakova, V. V. Dolgo-Saburova, et al., *Khim.-Farm. Zh.*, **31**(6), 19 – 22 (1997).
13. V. P. Krivonogov, G. A. Tolstikov, Yu. I. Murinov, et al., *Khim.-Farm. Zh.*, **31**(6), 23 – 27 (1997).
14. R. N. Galeeva, G. N. Novoselova, M. Yu. Gavrilov, et al., *Khim.-Farm. Zh.*, **31**(6), 38 – 39 (1997).
15. T. A. Smirnova, M. Yu. Gavrilov, L. P. Drovosekova, et al., *Khim.-Farm. Zh.*, **31**(3), 28 – 30 (1997).
16. Yu. Ya. Ivanov and N. M. Smirnova, *Khim.-Farm. Zh.*, **31**(1), 14 – 16 (1997).
17. B. Ya. Syropyatov, A. A. Gorbunov, V. S. Shklyaev, et al., *Khim.-Farm. Zh.*, **30**(11), 13 – 14 (1996).
18. V. A. Anisimova, A. A. Spasov, I. A. Bocharova, et al., *Khim.-Farm. Zh.*, **30**(1), 22 – 25 (1996).
19. J. Prous, *The Year's Drug News. Therapeutic Targets*, Prous Sci. Publ., Barcelona (1995).
20. V. V. Poroikov, D. A. Filimonov, and A. P. Boudunova, *Abstracts of Papers. The XIVth Int. Symp. on Medicinal Chemistry*, Maastricht (Netherlands) (1996), P-3.05.