Prediction of Biological Activity Spectra for Substances: Evaluation on the Diverse Sets of Drug-Like Structures

A.V. Stepanchikova*, A.A. Lagunin, D.A. Filimonov and V.V. Poroikov

Institute of Biomedical Chemistry RAMS, Pogodinskaya Str., 10, Moscow, 119121, Russia

Abstract: The concept of Biological Activity Spectrum served as a basis for developing PASS (Prediction of Activity Spectra for Substances) software product. PASS predicts simultaneously more than 780 pharmacological effects and biochemical mechanisms based on the structural formula of a substance. It may be used for finding new targets (mechanisms) for known pharmaceuticals and for searching new biologically active substances. PASS prediction ability was evaluated by activity spectra prediction for 63 substances that are presented in the Molecule of the Month section of Prous Science (http://www.prous.com), belong to different chemical classes and reveal various types of biological activity. Mean accuracy of prediction turned out to be about 90%; therefore, it is reasonable to use PASS for finding and optimizing new lead compounds. A web-site with a new internet version of PASS is introduced into practice in December 2001 (http://www.ibmh.msk.su/PASS). On the site, one can find a detailed description of the PASS approach as well as some examples of its applications, and estimate the quality of prediction by submitting structures of substances with known activities.

INTRODUCTION

The finding and optimisation of new lead compounds is the most important task of pharmaceutical R&D process.

SAR/QSAR/QSPR/Molecular Modelling methods are widely used for this purpose. Most of such methods of analysis (for review see: [1-3]) are very effective for predicting one or several types of biological activity within the same chemical series but are unable to elucidate the general biological "potential" of a molecule under study.

Here we represent PASS software product, that is able to elucidate many different possible types of biological activity of substances even if these substances do not belong to the same chemical classes. PASS predicts simultaneously pharmacological effects, mechanisms of action, mutagenicity, carcinogenicity, teratogenicity and embryotoxicity based exclusively on the structural formula of a substance [4].

Current release of the program (PASS 1.603) represents a significant step in PASS development. It provides prediction of 783 types of biological activity with the mean prediction accuracy in leave-one-out cross-validation (LOO CV) being 85% , while in 1998 [5] PASS could only predict 541 and in 1996 114 activity types [6] (incidentally, the mean prediction accuracy in 1996 was 78%). Comparing these results it is necessary to remember what the mean prediction accuracy is. Its value is 100% minus arithmetic average of the maximum error of prediction for each activity type. Taking into consideration the six-times increase in the number of predicted activity types, the essential increase in the prediction accuracy could be asserted. Such a tangible result was reaped from several fields.

Firstly, new MNA descriptors were developed for chemical structure representation [7]. Secondly, the number of compounds in the training set increased up to 45466 substances (cf. 30537 in 1998, and 9314 in 1996). In itself, the large number of substances in the training set is not enough for providing high quality of prediction. Only an advanced training set, containing reliable data on structure-activity relationships can seriously enhance the results of prediction. So the PASS training set was carefully prepared: biological activity of each substance was verified by several sources. Thirdly, the mathematical algorithm was changed to provide a higher accuracy and robustness of prediction.

As was already mentioned the mean accuracy of prediction is about 85% in leave-one-out cross validation. The LOO CV method is known to be very prone to overestimating the predictive accuracy unless the training set is truly diverse within the descriptor space. Therefore, we suggest that it is necessary to evaluate PASS on a diverse set of substances which belong to various chemical classes and reveal various types of biological activity. The set of substances from "Molecule of Month section of PROUS Science" [8] suits this purpose completely. In the present study we evaluate PASS prediction ability thereby examining the possibility of its use for searching new biologically active substances and new applications of well-known pharmaceuticals.

Biological Activity

We have used the "Biological Activity Spectrum of a substance" concept which seems to be a fundamental term

*Address correspondence to this author at the Institute of Biomedical Chemistry RAMS, Pogodinskaya Str., 10, Moscow, 119121, Russia; Phone: (7-095) 247-3029; E-mail: alla@ibmh.msk.su

0929-8673/03 $41.00+.00

for the description of biologically active substances. The "biological activity spectrum of a substance" is the list of biological activity names which reflects the result of chemical substance's interaction with different biological entities. We define the biological activity qualitatively (in "yes"/"none" terms) suggesting that the "biological activity spectrum" reflects the "intrinsic" property of a substance depending only on its structure and physical-chemical characteristics. At the same time the revealing of biological activity depends on the presence of the corresponding targets and conditions (experimental object, dose, route of administration, etc.).

The existing list of activity types contains both the names of pharmaco-therapeutic effects and the names of action mechanisms. In PASS all types of biological activity are considered as independent from one another, which is obviously an approximation because there exist some relationships between different activity types. However, this is a reasonable approximation used to obtain the probability estimates for different activity types and compounds from various chemical classes (see below).

METHODS

Our goal consists in providing maximally full information on the biological activity for any substance including such types of biological activity whose molecular mechanism of action is still unknown. This aim determines the choice of chemical substance representation and the mathematical method on which the provided algorithm is based.

The PASS approach is based on the suggestion, *Activity = Function (Structure)*. Thus, "comparing" structure of a new substance with that of well-known biologically active substances, it is possible to find out whether a new substance has a particular effect.

2D structural formula of a substance is chosen as the basis for structure description. Structure descriptors designated by our group, "Multilevel Neighborhoods of Atoms" (MNA), were proposed for chemical structure representation. The detailed description of MNA descriptors had been discussed earlier [7]. MNA descriptors simplify the structure representation in a manner that does not specify the bond types but includes hydrogen atoms according to the valence and partial charge of the atoms. MNA descriptors are generated as a recursively defined sequence:

- Zero-level MNA descriptor for each atom is the mark ***A*** of the atom itself;

- Any next-level MNA descriptor for each atom is the substructure notation $A(D_1D_2\ldots D_i\ldots)$, where D_i is the previous-level MNA descriptor for the ***i***-th immediate neighbors of the atom.

This iterative process can be continued including 2nd, 3rd, etc. neighborhoods of each atom. It is important to emphasize that the atom mark may include not only the atom type but also any additional information about the atom, for example, its belonging to a certain ring or chain in this PASS version. A structure of a molecule is represented in PASS as a set of the 1st- and 2nd-level MNA descriptors. In the 2nd-level MNA descriptors we used the mark "-" to define the belonging to a chain. Special computer experiments show that 3rd level of MNA descriptors is too specific. It occurs only in one structure among 50000.

In the case of long linear chains, level 2 may be insufficient. However, in the PASS training set consisting of drug-like substances such structures are not typical and their number is relatively small. Furthermore, in most cases heteroatoms or side groups are present in linear chains. See the examples of MNA descriptors in publication [7] or on our web-site [12].

The same set of descriptors is applied to the analysis and prediction of all types of biological activity. MNA descriptors could also be used to predict quantitative properties of a substance such as boiling point or mutagenicity [7].

PASS estimates the probabilities of a particular substance's belonging to the active and inactive sub-sets of a substance from the SAR Base (Structure-Activity Relationships Base). It is a complex knowledge base, containing vocabularies of MNA descriptors and activity names, the database of the substance structures presented by MNA descriptors, their biological activity types and data on SAR. Unfortunately, it is now impossible to collect sufficiently large number of active substances for all PASS activity types using available sources, that is why some activity types are presented by more than thousand substances, while some others are only represented by a few ones (see TABLE 1).

The substances are considered equivalent in PASS if they have the same molecular formula and the same MNA descriptors set. Since MNA descriptors do not represent the stereochemical peculiarities of a molecule, the substances, which only have stereochemical differences in the structure, are formally considered equivalent.

Algorithm of prediction is described in detail in [11].

The result of prediction is returned in the form of a table containing the list of biological activity names with the appropriate probability values – i.e. the values defining the likelihood for a given activity type to be either revealed (**Pa**) or not revealed (**Pi),** for each activity type from the predicted biological activity spectrum. Their values vary from 0.000 to 1.000. Only those activity types are considered possible for which **Pa** > **Pi**.

External Files of Substances

PASS uses SDfile or MOLfile formats [9] as an external source of structure and activity data to prepare both SAR Base and the set of substances to be predicted. SDfiles (*.sdf) can be exported either from ISIS/Base 2.0+ (MDL Information Systems, Inc.) or from another molecular editor which has the option of SD file's export. MOLfiles can be prepared by ISIS/Draw.

Training Set

The perfect training set should contain all substances ever tested for any biological activity, both active and inactive ones. Actually, the data on inactive substances is not always published in literature like any negative results of research. Thus, the current PASS training set is based mainly on the basis of "positive" information about the biological activity of chemical substances. If the data on a certain activity type for a particular substance was not found in the literature, the substance was considered "inactive" towards this type of biological activity. Obviously, this suggestion is overs-implified and may be wrong in some cases because if such data were not found in the literature, it means that either the substance was inactive or was not tested for this activity type. However, it was shown in special experiments that incompleteness of the data on substances from the training set does not significantly influence the results of prediction due to the robustness of approach used in PASS [10].

With PASS, the user can open the existing SAR Base and add new data to the SAR Base, create new SAR Base, execute the training procedure and select a sub-set of activity types for prediction.

Prediction Results

The Pa and Pi values vary from 0.000 to 1.000. To define the threshold for selecting types of activity to be predicted, the cutoff value of Pa should be chosen. Only activities with Pa value greater than the chosen threshold will be given in predicted activity spectra.

The quality of predictions is the main criterion of the program efficiency. The maximal error of prediction for each type of activity is shown in TABLE 1. In the "Prediction Results" window the user also obtains the total number of chemical descriptors for the substance. Reported is the number of descriptors which are new compared with the PASS training set descriptors.

Table 1. SAR Base Information

45466 Substances (the number of substances in the training set)

1482 Activity Types (the number of activity types in the training set)

41576 Descriptors (the number of descriptors in the training set)

783 Selected Activity Types (the number of selected activity types)

3 Minimum of substances, which must be contained in the training set with a particular type of activity for including to include this type of activity in the selected activities list.

14.908 Average MEP, % (Maximum error of prediction for each type of activity).

The list of Prous declared types of activity predicted by PASS

#*	Number of substances**	MEP, %***	Activity names
3	181	7.481	5 Alpha reductase inhibitor
8	22	7.490	5Hydroxytryptamine 1B agonist
9	131	5.794	5 Hydroxytryptamine 1D agonist
19	242	6.559	5 Hydroxytryptamine 3 antagonist
20	69	5.906	5 Hydroxytryptamine 4 agonist
38	47	8.951	Acetylcholine M3 receptor antagonist
42	498	14.645	Acetylcholine muscarinic antagonist
47	256	12.961	Acetylcholinesterase inhibitor
85	24	34.788	Alzheimer's disease treatment
91	895	22.695	Analgesic, non-opioid
99	169	28.193	Angiogenesis inhibitor
104	212	7.320	Angiotensin converting enzyme inhibitor
111	2011	22.099	Antiallergic
116	1405	26.232	Antiarthritic
117	659	25.621	Antiasthmatic
118	4268	15.921	Antibacterial
119	2621	21.379	Antihypercholesterolemic
161	4298	22.593	Antihypertensive
165	2715	24.740	Antiinflammatory
170	33	33.601	Antileishmanial
174	370	17.736	Antimigraine
180	5511	21.662	Antineoplastic
184	130	20.791	Antineoplastic enhancer
189	408	20.062	Antiosteoporotic
196	754	17.404	Antiprotozoal
203	514	23.256	Antipsoriatic
212	517	22.313	Antithrombotic
217	243	17.732	Antituberculosic
222	2357	21.766	Antiviral
223	3	2.188	Antiviral (CMV)
224	1130	17.821	Antiviral (HIV)
225	9	4.208	Antiviral (hepatitis B)
228	81	26.439	Antiviral (influenza)
260	523	16.986	Bronchodilator
299	9	7.438	Cholesterol absorption inhibitor
321	155	7.635	Cyclooxygenase 2 inhibitor
339	7	22.802	Dihydroorotate dehydrogenase inhibitor
346	61	9.932	Dopamine D1 antagonist
347	88	6.527	Dopamine D2 agonist

(Table. 1)contd.....

(Table. 1)contd.....

#*	Number of substances**	MEP, %***	Activity names
367	126	7.730	Endothelin A receptor antagonist
368	62	9.767	Endothelin B receptor antagonist
370	271	7.893	Endothelin receptor antagonist
383	147	6.536	Factor Xa inhibitor
384	203	10.682	Farnesyltransferase inhibitor
400	361	17.032	Gastric antisecretory
451	30	26.545	HDL-cholesterol increasing
454	208	5.510	HMG CoA reductase inhibitor
484	1179	17.078	Hypolipemic
489	828	27.081	Immunosuppressant
493	23	27.880	Insulin promoter
503	150	23.015	Irritable Bowel syndrome treatment
523	639	12.135	Lipoxygenase inhibitor
534	69	29.521	Male reproductive disfunction treatment
537	105	9.477	Matrix metalloproteinase inhibitor
547	181	15.533	Microtubule formation inhibitor
561	414	11.049	NMDA receptor antagonist
569	35	4.588	Neuraminidase (influenza) inhibitor
580	136	3.138	Neutral endopeptidase inhibitor
584	282	18.079	Non-steroidal antiinflammatory agent
615	43	15.064	Phosphodiesterase V inhibitor
626	1637	18.592	Platelet aggregation inhibitor
652	6	12.808	Prostate cancer treatment
653	313	15.455	Prostate disorders treatment
694	88	2.349	Retinoid acid receptor agonist
695	183	14.576	Reverse transcriptase inhibitor
756	13	11.776	Tryptase inhibitor
766	168	18.611	Urinary incontinence treatment
771	648	22.243	Vasodilator
779	114	0.712	Vitamin D-like

*The identifier of this particular type of activity in PASS training set; **The number of substances in PASS training set revealing this particular type of activity; ***MEP is error of prediction of this particular type of activity obtained by Leave-One-Out Cross-Validation method.

TESTING PASS ON PROUS "MOLECULE OF MONTH" SET OF SUBSTANCES

Each month "Molecule of Month" section of Prous Science web-site highlights [8] different drug molecule or molecules. Selection is based on the following criteria:

- The originality of the chemical structure;
- The singularity of the mechanism of action;
- The drug's progression through the R&D pipeline;
- Its use in a new indication or where current therapies are inexistent or have proved unsatisfactory.

80 substances were represented by 01.04.2002. We chose only those low-molecular weight substances whose structural formulae were presented. These are the so-called drug-like substances. Thus, 63 substances were chosen for an evaluation set.

Our sub-set contains 63 substances from Prous "Molecule of Month" set. These diverse molecules do not belong to the same chemical series. They are derivatives of benzimidazole, imidazole, indole, thiazolidine, thiazepine, carbazole, furan, naphthyridine, quinoline, isoxazole, purine, pyrimidine, naphthalene, pyrrole, pyrazole, piperidine, benzoxazine, benzothiophene, benzoic acid, pyridine, thiochromane, pyrazolopyrimidine, and piperazine.

PROUS reports two or three types of biological activity for each substance. PASS generates many activity types which can be revealed by the substance. The appropriately calculated probability values Pa and Pi for every type of activity to be revealed are also reported. Comparison of predicted results with experimental ones provides the data for estimating PASS predictive power. Information on supplementary effects can also be very important, because the results of prediction, if confirmed experimentally, may become a reason for a new application of these pharmaceuticals.

Biological Activity Spectra was predicted for all these substances. Prediction was carried out in the mode of excluding the equivalent substances. In other words, substances equivalent to those under prediction are excluded from the PASS training set. It turned out that 51 of 63 substances are included in the PASS training set and each of them was removed from the PASS training set during the prediction. Five out of the remaining have no new descriptor, two substances have one new descriptor, two have two new ones, and one substance has three new ones. Prediction results are presented in TABLE 2.

The names of all 63 substances and the corresponding names of biological activities reported by PROUS for each substance are presented herein. Corresponding values of Pa and Pi for each type of activity are also reported. Among 159 types of activity (i.e. the sum total for all types of activities throughout all 63 substances) declared by PROUS, 15 types would not be predicted for one of the three reasons: either this type of activity is absent in the PASS activity list, or this type of activity is presented in the PASS training set by less than 3 substances, or else this type of activity is not used in the prediction because the MEP value is too large. 130 types of activity were predicted by PASS, and 14 were not predicted. Thus, the average accuracy of prediction is 90%. A natural question arizes: how good the prediction result is? Predicted activities of 53 substances from these 63 substances completely coincided with PROUS declared types of activity. None of the biological activity types declared by PROUS was confirmed for 2 of the substances (#20 - APC-366, #63 - Sch-351125). Rho kinase

inhibitor was not predicted for Y-27632 (#53) because this type of activity was not presented in the PASS activity list. This is an encouraging result demonstrating that PASS is a reasonably reliable instrument for pilot analysis of biological potential of any drug-like substance which is very important for planning experiments in medicinal chemistry. Comprehension and interpretation of prediction results is an essential part of work. If a set of MNA descriptors of a substance contains one or more new descriptors, it means that this substance is really a new one, compared with those in the training set. Therefore, such a substance should attract the researcher's attention. There are five such substances in the PROUS set. Practically, all types of biological activity declared by PROUS are confirmed by PASS prediction (see TABLE 2).

Table 2. Prediction Results

Pharmaceutical	PROUS declared types of activity	Pa	Pi	Comments
1Esomeprazole magnezium	Gastric antisecretory	0.492	0.007	1 new descriptor
2 AG-7088	3C protease (rhinovirus) inhibitor* Antiviral	– 0.426	– 0.040	 +
3 Quinupristin	Antibiotic	0.446	0.011	+
4 Dalfopristin	Antibiotic (*Antibacterial*)	0.688	0.007	+
5 Alosetron hydrochloride	5 Hydroxytryptamine 3 antagonist Irritable Bowel syndrome therapy	0.992 0.992	0.003 0.003	+
6 Zanamivir	Neuraminidase (influenza) inhibitor Antiviral (influenza)	0.907 0.907	0.000 0.020	+
7 Tegaserod maleate	Irritable Bowel syndrome therapy 5 Hydroxytryptamine 4 agonist	0.522 –	0.008 –	 +
8 Rosiglitazone maleate	Antidiabetic	0.906	0.001	+
9 Omapatrilat	Antihypertensive Neutral endopeptidase inhibitor Angiotensin converting enzyme inhibitor	0.958 0.958 0.946	0.002 0.002 0.002	 +
10 Frovatriptan	Antimigraine 5 Hydroxytryptamine 1B agonist 5 Hydroxytryptamine 1D agonist	0.579 0.261 0.422	0.005 0.006 0.006	 +
11 Verteporfin	Antipsoriatic Antileukemic (*Antineoplastic*) Antiarthritic	0.479 0.932 0.329	0.048 0.005 0.223	 +
12 Bay-12-9566	Antiarthritic Matrix metalloproteinase inhibitor Antineoplastic Angiogenesis inhibitor	0.826 0.678 0.678 0.574	0.004 0.004 0.004 0.039	 +
13 Rofecoxib	Cyclooxygenase 2 inhibitor Antiinflammatory Antiarthritic (Rheumatoid arthritis treatment)	0.889 0.936 0.936	0.002 0.002 0.002	 +
14 SB-265805	Antibacterial	0.768	0.001	0 new descriptor
15 Moxifloxacin Hydrochloride	Antibacterial Antibiotic Antituberculosic	0.768 0.768 –	0.001 0.001 –	 +
16 Leflunomide	Antiarthritic Dihydroorotate dehydrogenase inhibitor Antineoplastic Tirosine kinase inhibitor Rheumatoid arthritis treatment* Epidermal growth factor antagonist*	0.655 0.342 – –	0.009 0.019 – –	 +
17 Abacavir succinate	Antiviral (HIV) Reverse transcriptase inhibitor	0.600 0.427	0.008 0.008	+
18 Capecitabine	Antineoplastic	0.850	0.005	+

(Table 2) contd.....

Pharmaceutical	PROUS declared types of activity	Pa	Pi	Comments
19 Paricalcitol	Vitamin D-like Parathyroid hormone antagonist (ПерестаП)	0.761 -	0.001 -	 +
20 APC-366 (Arris; Bayer)	Tryptase inhibitor Antiasthmatic	– –	– –	 +
21 Marimastat (British Biotech)	Matrix metalloproteinase inhibitor Antineoplastic Prostate cancer treatment	0.983 0.955 0.131	0.001 0.002 0.060	 +
22 Sildenafil (Viagra)	Male reproductive disfunction treatment Phosphodiesterase V inhibitor Erective disfunction treatment**	0.962 0.950	0.002 0.002	 +
23 NN-42-1007 (Novo Nordisk)	Antidiabetic Glycogen phosphorylase inhibitor*	0.309	0.049	+
24 P-52608	Antiarthritic Antiinflammatory Retinoid acid receptor agonist**	0.919 0.221	0.006 0.069	 +
25 Clopidogrel hydrogensulfate	Platelet aggregation inhibitor Antithrombotic	0.476 0.685	0.022 0.005	 +
26 Celecoxib	Cyclooxygenase 2 inhibitor Antiinflammatory Rheumatoid arthritis treatment (Antiarthritic)	0.773 0.917 0.917	0.002 0.002 0.002	 +
27 Donepezil hydrochloride	Acetylcholinesterase inhibitor Alzheimer's disease treatment	0.775 0.775	0.005 0.005	 +
28 Efavirenz	Reverse transcriptase inhibitor Antiviral (HIV)	0.981 0.823	0.002 0.006	+
29 epothilone A (EN: 222566)	Microtubule formation inhibitor Antineoplastic	0.471 0.787	0.004 0.004	+
30 epothilone B	Microtubule formation inhibitor Antineoplastic	0.406 0.820	0.006 0.001	+
31 epothilone analogues	Microtubule formation inhibitor Antineoplastic	0.791 0.791	0.001 0.001	0 new descriptor
32 GS-4104	Neuraminidase (influenza) inhibitor Antiviral (influenza)	0.466 0.505	0.001 0.001	 +
33 HMR-3647	Antibacterial Antibiotic	0.714 0.718	0.005 0.001	+
34 Raloxifene hydrochloride	Antiosteoporotic Antineoplastic Cholesterol lowering**	0.839 0.683	0.005 0.005	 +
35 LGD-1069	Antineoplastic Retinoid acid receptor agonist Antidiabetic Insulin promoter	0.709 0.709 0.378 0.365	0.002 0.002 0.005 0.104	 +
36 Lobucavir	Antiviral Antiviral (HIV) Antiviral (CMV) Antiviral (hepatitis B) Antiviral (herpes)	0.925 0.620 0.074 0.611 0.646	0.005 0.007 0.005 0.007 0.005	 +
37 Atorvastatin	HMG CoA reductase inhibitor Hypolipemic	0.625 0.851	0.002 0.002	+
38 Repaglinide	Insulin promoter Antidiabetic	0.506 0.439	0.041 0.076	+
39 Ro-61-1790	Endothelin A receptor antagonist Antihypertensive Anticerebroischemic Vasodilator	0.757 0.940 – 0.917	0.003 0.002 – 0.002	 +

(Table 2) contd.....

Pharmaceutical	PROUS declared types of activity	Pa	Pi	Comments
40 Tazarotene	Antipsoriatic Retinoid acid receptor agonist	0.963 0.963	0.001 0.001	+
41 VX-710	Antineoplastic Antineoplastic enhancer	0.724 0.724	0.006 0.006	2 new descriptor
42 Zileuton	Lipoxygenase inhibitor Antiinflammatory Antiallergic Antiasthmatic	0.988 0.988 0.991 0.988	0.002 0.002 0.002 0.002	+
43 STI-571	Tirosine kinase inhibitor Antileukemic***	0.426	0.007	+
44 Sitaxsentan sodium	Endothelin A receptor antagonist Heart failure treatment* Antihypertensive Pulmonary hypertension treatment	0.924 0.975 –	0.002 0.001 –	0 new descriptor
45 Darusentan (LU-135252)	Endothelin A receptor antagonist Vasodilator Vasodilator, coronary	0.952 0.952	0.002 0.002	+
46 Tezosentan disodium (Ro-61-0612)	Endothelin A receptor antagonist Endothelin B receptor antagonist	0.854 0.707	0.002 0.004	0 new descriptor
47 J-104132 (L-753037)	Endothelin A receptor antagonist Endothelin B receptor antagonist	0.613 0.450	0.004 0.006	+
48 Tiotropium bromide	Acetylcholine muscarinic antagonist Acetylcholine M1 receptor antagonist Acetylcholine M2 receptor antagonist Acetylcholine M3 receptor antagonist Bronchodilator	0.979 - - 0.809 0.960	0.003 - - 0.002 0.003	+
49 HCT-1026	Non-steroidal antiinflammatory agent Antithrombotic Antiosteoporotic Urinary incontinence treatment	0.929 0.533 0.480 0.220	0.005 0.079 0.097 0.031	+
50 SDZ-GLC-756	Antiglaucomic Dopamine D2 agonist Dopamine D1 antagonist	0.227 0.645 0.267	0.075 0.004 0.017	+
51 Dutasteride	5 Alpha reductase inhibitor Prostatic (benign) hyperplasia therapy*** Prostate disorders treatment	0.937 0.937	0.002 0.002	0 new descriptor
52 V-Glycopeptide (Bl-397)	Antibiotic Antibiotic Glycopeptide-like	0.672 0.672	0.001 0.001	0 new descriptor
53 Y-27632	Rho kinase inhibitor*			+
54 Fondaparin Sodium	Antithrombotic Factor Xa inhibitor	0.959 0.907	0.005 0.003	+
55 R-115777	Farnesyltransferaseinhibitor Antineoplastic	0.940 0.961	0.002 0.002	+
56 Miltefosine	Antileishmanial Antiprotozoal	- 0.873	- 0.002	+
57 Atrasentan	Endothelin receptor antagonist Antihypertensive Endothelin A receptor antagonist Antiasthmatic Prostate cancer treatment Pulmonary hypertension treatment* Heart failure treatment** Erective disfunction treatment**	0.760 0.895 0.579 0.378 -	0.004 0.003 0.004 0.014 -	+
58 Rosuvastatin calcium	HMG CoA reductase inhibitor HDL-cholesterol increasing	0.620 0.339	0.002 0.165	+

(Table 2) contd.....

Pharmaceutical	PROUS declared types of activity	Pa	Pi	Comments
59 Flupirtine	Analgesic, non-opioid NMDA receptor antagonist Apoptosis antagonist*	0.629 0.158	0.017 0.054	0 new descriptor
60 Ezetimibe	Cholesterol absorption inhibitor Antihypercholesterolemic	0.179 0.250	0.007 0.024	+
61 ACE2 inhibitor	Antihypertensive Angiotensin converting enzyme inhibitor Heart failure treatment **	0.593 0.137	0.012 0.021	0 new descriptor
62 Pimecrolimus	Antiinflammatory Immunosuppressant	0.992 0.992	0.002 0.002	0 new descriptor
63 Sch-351125	Antiviral (HIV)	–	–	1 new descriptor

♦ "+ " in "comments" column marks those substances, which belong to PASS training set. If a substance has new descriptors, which were not found in PASS descriptor vocabulary, the number of these descriptors is presented here.

* This type of activity is absent from PASS activity list.

** This type of activity is presented in PASS training set by less than 3 substances (see TABLE 1).

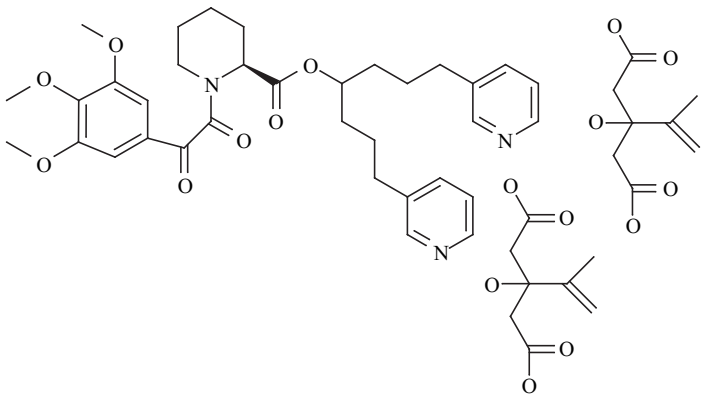
*** This type of activity is not used in the prediction because of the MEP value is too great (see TABLE 1).

An example of biological activity spectra for **VX-710** is given in Fig. (**1**). One can see that both types of biological activity declared by PROUS were predicted. Besides, PASS predicts several new types of biological activity for this substance. It is necessary to mention the relative novelty of this substance for PASS. It is not only absent in the PASS training set but also has two new MNA descriptors that are absent in the MNA vocabulary. In spite of that, both types of biological activity declared by PROUS were predicted with high probability. Let us now analyze this example in detail. The substance consists of three parts: basic part of the molecule and 2 molecules of 3-hydroxy-3-1isopropenyl-pentanedionic acid. The latter two do not influence the biological activity (according to the PASS prediction), but generate two descriptors that are new relating to the compounds from the PASS training set. Prediction of the biological activity spectra for the basic part of the molecule gives Pa = 0.9503 for Antineoplastic enhancer.

On the other hand, in some cases a predicted biological activity spectra may include some side- and toxic effects that provide a reason for eliminating such substances from R & D pipeline. Therefore, our experiments with the PROUS "Molecule of Month" subset demonstrated that PASS prediction ability is significant enough to use the program for finding new biological activity types in known drugs [11].

In addition, it is necessary to mention a new Internet version of the program (PASS Inet) introduced in December 2001 [12]. It is a www server for the on-line prediction of the Biological Activity Spectra for Substances [13], which

42 VX-710



71 Substructure descriptors; 2 new.
Possible activities at Pa > 0.4.

Pa	Pi	for Activity:
0.724	**0.006**	**Antineoplastic enhancer**
0.724	**0.006**	**Antineoplastic**
0.630	0.055	Fibrinolytic
0.582	0.015	Neurotrophic factor
0.544	0.019	Radiosensitizer
0.533	0.076	Antacid
0.582	0.015	Cognition disorders treatment
0.458	0.113	Immunosuppressant
0.582	0.015	Nootropic
0.430	0.105	Cholesterol synthesis inhibitor
0.472	0.137	Alzheimer's disease treatment
0.459	0.146	Acetylcholine release stimulant
0.434	0.129	Cytokine modulator
0.458	0.129	Immunomodulator
0.460	0.166	Lipid metabolism regulator
0.445	0.127	Interleukin 1 antagonist
0.422	0.140	Interleukin antagonist
0.434	0.164	Convulsant
0.469	0.063	Multiple sclerosis treatment
0.445	0.063	Septic shock treatment

Fig. (1). Part of predicted activity spectra for VX-710.

Only those types of activity for which Pa > 0.4 are shown. PROUS declared (i.e. experimentally confirmed) activity types are marked in bold.

contains about 45466 biologically active substances in a training set and predicts biological activity spectra for 783 types of pharmacological effects, action mechanisms, mutagenicity, carcinogenicity, teratogenicity and embryotoxicity. To get the Biological Activity Spectrum of the substance, one should send a standard MOLfile via Internet. MOLfile can be prepared by using the ISIS/Draw chemical editor [8]; or the structural formula of a substance can be drawn using Marvin applet [14]. There one can find a detailed information on the algorithm, applications and efficiency of PASS. Besides, one has an opportunity to estimate the accuracy of biological activity spectra prediction by submitting substances with known activities.

REFERENCES AND NOTES

[1] Kubinyi, H., Folkers, G., Martin, Y.C. Eds.; *3D QSAR in Drug Design: Recent Advances;* Kluwer/Escom, **1998**.

[2] Lipnick, R.L. *SAR & QSAR Environ. Res.*, **1999**, *10* (2-3), 239-248.

[3] Ed., B. Testa. *Pharmacokinetic Optimization in Drug Research*; Wiley-VCH: New York **2001**.

[4] Poroikov, V.V.; Filimonov, D. Abstr. XVIth Intern. Symp. Medicinal Chemistry, Bologna (Italy), **2000**; p.149.

[5] Gloriozova, T.A.; Filimonov, D.A.; Lagunin, A.A.; Poroikov, V.V. *Chim. Pharm. J. (Rus).* **1998**,.*32* (12), 32-39. (English translation by Consultants Bureau, New York: Pharmaceutical Chemistry Journal, **1998**, *32* (12), 658-664.

[6] Poroikov, V.V.; Filimonov, D.A. In: QSAR and Molecular Modelling Concepts, Computational Tools and Biological Applications. Barcelona: Prous Science Publishers; **1996**; p. 49-50.

[7] Filimonov, D.A; Poroikov, V.V; Borodina, Y.; Gloriozova, T. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 666-670.

[8] [http://www.prous.com/mom/]

[9] [http://www.mdli.com]

[10] Poroikov, V.V.; Filimonov, D.A.; Borodina, Yu. V.; Lagunin, A.A.; Kos, A. *J. Chem. Inform. Comput. Sci.*, **2000**, *40* (6), 1349-1355.

[11] Poroikov, V.V.; Akimov, D.; Shabelnikova, E.; Filimonov, D. *SAR and QSAR in Environmental Research*. **2001,** *12* (4), 327-344.

[12] [http://www.ibmh.msk.su/PASS/]

[13] Lagunin, A.; Stepanchikova, A.; Filimonov, D.; Poroikov, V.**V.** *Bioinformatics,* **2000**, *16* (8), 747-748.

[14] [ChemAxon Ltd. http://www.chemaxon.com]