# COMPUTER-AIDED PREDICTION OF BIOLOGICAL ACTIVITY SPECTRA. APPLICATION FOR FINDING AND OPTIMIZATION OF NEW LEADS

*V. Poroikov, D. Filimonov*

Institute of Biomedical Chemistry of Russian Academy of Medical Sciences; 10, Pogodinskaya Street, Moscow, 199832, Russia; E-mail: vvp@ibmh.msk.su

## INTRODUCTION

Each biologically active compound possesses a number of biological activities. Its specificity of action is always relative and is defined by the peculiarities of object, dose, route, etc. On the contrary, the biological potential of compound includes all activities, which can be discovered under some specific experimental conditions. We call such biological potential the biological activity spectrum. Biological activity spectrum of compound can be predicted on the basis of structure-activity relationships found by the analysis of the known data from the training set. Based on the analysis of large training set consisting of tens of thousands known biologically active compounds, computer program PASS (1, 2, 3) provides the means to evaluate any new compound in huge chemical-pharmacological space.

## BACKGROUND

Computer program PASS is the product of ideas originated more than 25 years ago within the framework of the National Registration System of New Chemical Compounds organized in the USSR in 1972 (4). It was V.Avidon who suggested that many kinds of biological activity could be predicted on the basis of structural formulae of chemical compounds (5). Similar approach was under development by V.Golender and A.Rozenblit (6).

In 1985, when the authors joined in the project, previous experience was taken into account. The only published example of a comparable approach developed and applied outside the Soviet Union was the U.S. National Cancer Institute's Drug Information System (7). In that time cooperation with scientists from Western countries was restricted and our team had to follow its own way. It was the way of trial and error, but due to this the team developed an innovative concept of biological activity spectra according to which the compound's biological action is considered as an intrinsic property of the structure (8), an original mathematical algorithm (9), new universal chemical descriptors (10), new training

set (3). The program, which predicts more than 500 kinds of biological activity, was created on this basis.

In the past years computer program PASS was reconstructed several times and its current version 1.41 is rather different even from the product which was available several years ago (8). During this evolution the accuracy and robustness of predictions has been increased substantially.

## BRIEF DESCRIPTION OF PASS

Computer system PASS (*Prediction of Activity Spectra for Substances*) predicts simultaneously several hundreds of biological activities (pharmacological main and side effects, mechanisms of action, mutagenicity, carcinogenicity, teratogenicity and embryotoxicity). The biological activity spectrum of a compound presents all compound's actions despite the difference in essential conditions of its experimental determination. If the difference in species, sex, age, dose, route, etc. is neglected, the biological activity can be identified only qualitatively. Thus, "the biological activity spectrum" is defined as the "intrinsic" property of a compound depending only on its structure and physico-chemical characteristics.

Prediction of this spectrum by PASS is based on SAR analysis of the training set containing more than 35,000 compounds which have more than 500 kinds of biological activity (3). Therefore, PASS once trained is able to predict simultaneously all biological activities which are included in the training set. To provide the best quality of prediction a new information about biologically active compounds is collected permanently from papers and electronic sources and, after the experts' evaluation, is regularly added to the training set.

Chemical descriptors used in PASS analysis, called multilevel neighborhoods of atoms (MNA), are recently considered in detail (10). They are automatically generated by on the basis of MOL-file (11) of a molecule. The list of MNA descriptors currently consists of more than 35700 different items. The new ones are added to this list being founded in a novel compound refreshing the training set. MNA descriptors are effectively applied in SAR, QSAR and similarity analysis (3, 10). They can be also used as keys or fingerprints to cluster the libraries of chemical compounds, to select the representative sub-sets from chemical databases, etc.

PASS algorithm was chosen by theoretical and empirical comparison of many different mathematical methods to provide high accuracy of prediction and robustness of calculated estimates (12). It was shown that the mean accuracy of prediction with PASS is about 86% in LOO cross-validation (3). Using MDDR database (11) to create heterogeneous training and evaluation sets it was recently demonstrated that the predictions are robust despite excluding up to 60% of information (13).

PASS uses MOL or SD-files (11) as input and the results of prediction (output) can be obtained as CSV, TXT or SD-files. Since the prediction of biological activity spectra for 1,000 compounds in usual PC takes about 1 minute, PASS can be effectively applied to predict biological potential of separate compounds and to analyze large chemical databases.

PASS testing in a blind mode by 9 scientists from 8 countries versus the heterogenous set of 118 compounds having 138 activities demonstrated the mean accuracy of prediction 82.6% (14).

## APPLICATIONS OF PASS

Example of prediction for well-known drug Buspiron is presented below. Pa and Pi are the estimates of probability to be active and inactive respectively. Their values vary from

zero to one. Only activities with Pa>0.5 are shown here. Known activities are marked in bold.

PASS 1.41 - Prediction of Activity Spectra for Substances
    Copyright (C) 1998-2000 V.V.Poroikov, D.A.Filimonov & Associates

Chemical Structure File: buspiron.mol
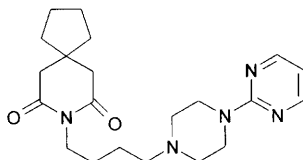


34 Substructure descriptors; 0 new.

| Pa | Pi | Activity: |
|---|---|---|
| **0.981** | **0.003** | **5 Hydroxytryptamine 1A agonist** |
| **0.969** | **0.006** | **Psychotropic** |
| **0.946** | **0.006** | **Anxiolytic** |
| **0.933** | **0.005** | **5 Hydroxytryptamine agonist** |
| **0.925** | **0.004** | **5 Hydroxytryptamine 1 agonist** |
| **0.851** | **0.006** | **Antipsychotic** |
| **0.836** | **0.007** | **Psychosexual dysfunction treatment** |
| **0.801** | **0.008** | **Antidepressant** |
| 0.709 | 0.007 | Alpha adrenoreceptor antagonist |
| 0.623 | 0.010 | Neurotrophic factor |
| 0.597 | 0.008 | Dopamine antagonist |
| 0.579 | 0.007 | Dopamine D2 antagonist |
| 0.589 | 0.037 | Rhinitis treatment |
| 0.571 | 0.020 | Cognition disorders treatment |
| 0.573 | 0.038 | Cytokine modulator |
| **0.566** | **0.044** | **Hypotermic** |
| 0.501 | 0.019 | Sodium channel blocker |

   Buspiron's structure has no MNA descriptor, which has not been found at least once in compounds from the training set. Most of known activities are predicted with high probabilities. Both main (Anxiolytic, etc.) and side (Hypotermic) effects are predicted. Some new effects (Rhinitis treatment, Cognition disorders treatment) and mechanisms (Alpha adrenoreceptor antagonist, Neurotrophic factor, etc.) are predicted too. This new knowledge can be used to determine the priorities in further Buspiron's investigation, which may help to find some new indications for this pharmaceutical and to avoid some adverse reactions in medical use.
   Such option of PASS application can be utilized to drug-candidates or even to well-known drugs, used for many years. For instance, Gedeon Richter recommended to use the results of biological activity spectra prediction as the basis for further investigation of Cavinton (15).
   PASS can be also applied for computer screening of chemical compounds with required activities. There exist many companies, which collect and resell the samples of chemical compounds, see for instance: (16, 17, 18). Hundreds of thousands compounds from stock are available for screening and millions of compounds can be synthesized on request. Most of chemical databases are arranged under the DBMS ISIS (11) exports the SD structural files. These files can be used as the input for PASS. The output will give the results of biological activity prediction, which can be further used for targeting or focusing of the library.

For instance, application of such approach to the CMC database (11), including about 7,000 biologically active compounds, reveals many compounds with predicted hypnotic effect (Figure 1). For 72 of them hypnotic effect has been confirmed by the experiment. Some other compounds from this focused collection were never tested versus hypnotic activity and may serve as a source for new leads (19).
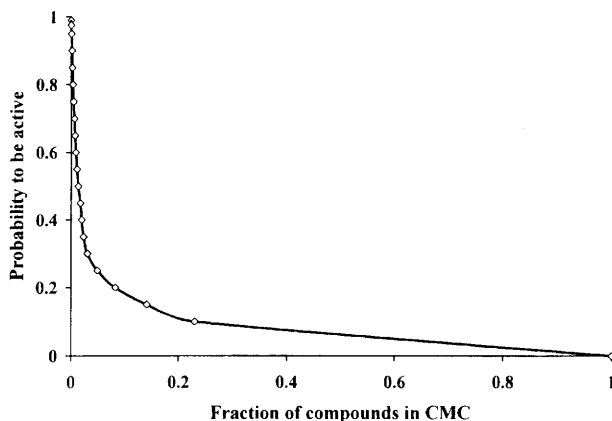


*Figure 1.* Distribution of fraction of compounds' with predicted hypnotic activity in CMC vs. the calculated values of Pa.

Usual interpretation of prediction results is based on the Pa values. If Pa > 0.7 the chance to find the activity in experiment is high, but in many cases the compound may occur to be the close analogue of known pharmaceutical agents. If 0.5 < Pa < 0.7 the chance to find the activity in experiment is less, but the compound is not so similar to known pharmaceutical agents. If Pa < 0.5 the chance to find the activity in experiment is even more less, but if it will be confirmed the compound might occur to be a New Chemical Entity. It was shown that the numbers of experiments necessary to find the known effects for the heterogenous set of substances by using PASS prediction and without it are 1650 and 316 respectively (20). Therefore, the economic viability of PASS application exceeds 500%.

On the basis of predicted biological activity spectra new lead structures were discovered with the following effects: antiamnestic (21); antiulzer (22), antitumor (23), antiinflammatory, antioxidant and anestetic (24), etc. – for more examples see (3).


## CONCLUSIONS

Computer program PASS can be effectively used for:

1. Finding of compounds with required properties and without undesirable side effects.
2. Revealing of new effects and mechanisms of action for known substances from the corporate and personal databases.
3. Selecting of the most prospective compounds from the set of available samples for highthroughput screening.
4. Determining of more relevant screens for particular compound.

# REFERENCES

1. Filimonov, D.A., Poroikov, V.V., Karaicheva ,E.I., Kazaryan, R.K., Boudunova, A.P., Mikhailovsky, E.M., Rudnitskih, A.V., Goncharenko, L.V., Burov, Yu.V. *Computer-aided prediction of biological activity spectra of chemical substances on the basis of their structural formulae: computerized system PASS*. Experimental and Clinical Pharmacology (Rus) 1995, 8(2): 56-62.
2. Filimonov, D. A., Poroikov, V.V. PASS: *Computerized prediction of biological activity spectra for chemical substances*. In: Bioactive Compound Design: Possibilities for Industrial Use, Oxford: BIOS Scientific Publishers 1996, 47-56.
3. http://www.ibmh.msk.su/PASS
4. Burov, Yu.V., Poroikov, V.V., Korolchenko, L.V. *National system for registration and biological testing of chemical compounds: facilities for new drugs search*. Bull. Natl. Center for Biologically Active Compounds (Rus.) 1990, 1: 4-25.
5. Avidon, V.V. *The criteria of chemical structures similarity and the principles for design of description language for chemical information processing of biologically active compounds*. Chim.-Pharm. J. (Rus) 1974, 8: 22-25.
6. Golender, V.E., Rosenblit, A.B. *Logical and Combinatorial Algorithms for Drug Design*. Wiley&Sons: L., a.o. 1983.
7. Milne, G.W.A., Miller, J.A. *Drug information system. 1. System overview*. J. Chem. Inf. Comput. Sci. 1986, 36 (2): 154-159.
8. Gloriozova, T. A., Filimonov, D. A., Lagunin, A. A., Poroikov, V. V. *Evaluation of computer system for prediction of biological activity PASS on the set of new chemical compounds*. Chim.-Pharm. J. (Rus) 1998, 32(12): 32-39.
9. Filimonov, D. A., Poroikov, V.V. *Robust prediction of many biological activities*. Abstr. 11th European Symposium on Quantitative Structure-Activity Relationships (September 1-3, Lausanne) 1996, Abstr. P-59A.
10. Filimonov, D., Poroikov, V., Borodina, Yu., Gloriozova, T. *Chemical similarity assessment through multilevel neighborhoods of atoms: definition and comparison with the other descriptors*. J. Chem. Inf. Comput. Sci. 1999, 39(4): 666-670.
11. http://www.mdli.com
12. Filimonov, D.A. *Comparison of different algorithms of computer-aided biological activity spectra prediction on the basis of structural formulae of chemical compounds*. II Rus. Natl. Congr. "Man and Drugs" (April 10-15, Moscow) 1995, 62-63.
13. Poroikov, V.V., Filimonov, D.A., Borodina, Yu.V., Lagunin, A.A., Kos, A. *Robustness of biological activity predicting by computer program PASS for non-congeneric sets of chemical compounds*. J. Chem. Inf. Comput. Sci. 2000, accepted for publication.
14. http://www.chemint.org/chemint99/abstracts/lectures.html
15. Poroikov, V.V., Filimonov, D.A., Gloriozova, T.A., Lagunin, A.A., Stepanchikova, A.V. *Predicted biological activity spectrum for Cavinton*. In: Investigation of Cavinton in experiment and clinical practice. Advisable approaches. Moscow: Gedeon Richter 1998, 42-45.
16. http://www.specs.net
17. http://www.chemdiv.com
18. http://www.chembridge.com/
19. Poroikov, V.V. *Computer aided prediction of biological activity spectra: exhaustive evaluation of chemistry-pharmacology space*. NATO Advanced Research Workshop "Frontiers in molecular diversity: from biology to material science" (18–23 September, Moscow) 1999, Abstr.28.
20. Poroikov, V.V, Filimonov, D.A, Boudunova, A.P. *Computer assisted prediction of biological activity spectra: estimating the effectivity of use in high throughput screening*. XIVth International Symposium on Medicinal Chemistry (8-12 September, Maastricht) 1996, Abstr: P-3.05.
21. Filimonov, D.A., Poroikov, V.V., Boudunova, A.P., Rudnitskih, A.V., Burov, Yu.V. *Computerized Prediction of Antiamnestic Activity for Chemical Compounds: PASS Possibilities Extending*. SCI Conference "Design of Bioactive Compounds" (4-7 December, Potsdam) 1995, Abstr. 26.
22. Trapkov, V.A., Budunova, A.P., Burova, O.A., Filimonov, D.A., Poroikov, V.V. *Discovery of New Antiulcer Agents by Computer Aided Prediction of Biological Activity*. Problems in Medical Chemistry (Moscow) 1997, 43(1): 41-57.
23. Pogrebnyak, A.V., Poroikov, V.V., Starych, V.V., Konovalov, D.A. *Computer prediction of anticancer activity of sesquiterpene lactones discovered in representatives of Asteracea*e. Plant Resources 1998, 34 (1): 61-64.
24. Geronikaki, A., Poroikov, V., Hadjipavlou-Litina, D., Filimonov, D., Lagunin, A., Mgonzo, R. *Computer aided predicting the biological activity spectra and experimental testing of new thiazole derivatives*. Quant. Struct.-Activ. Relationships 1999, 18(6): 630-639.